Cosmic Cartography of the Large-Scale Structure with Sloan Digital Sky Survey Data Release 6

Francisco S. Kitaura,[1,2][*] Jens Jasche,[2] Cheng Li,[2,3] Torsten A. Enßlin,[2] R. Benton Metcalf,[2] Benjamin D. Wandelt,[4] Gerard Lemson,[5,6] and Simon D. M. White,[2]

[1] *SISSA, Scuola Internazionale Superiore di Studi Avanzati, via Beirut 2-4 34014 Trieste, Italy*
[2] *MPA, Max-Planck Institut für Astrophysik, Karl-Schwarzschildstr. 1, D-85748 Garching, Germany*
[3] *MPA/SHAO Joint Center for Astrophysical Cosmology at Shanghai Astronomical Observatory, Nandan Road 80, Shanghai 200030, China*
[4] *Department of Physics University of Illinois at Urbana-Champaign, 1110 West Green Street Urbana, IL 61801-3080, USA*
[5] *Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstr. 12-14, 69120 Heidelberg, Germany*
[6] *MPE, Max-Planck Institut für Extraterrestrische Physik, Giessenbachstrae, D-85748 Garching, Germany*

6 November 2009

ABSTRACT
We present the largest Wiener reconstruction of the cosmic density field made to date. The reconstruction is based on the Sloan Digital Sky Survey data release 6 covering the northern Galactic cap. We use a novel supersampling algorithm to suppress aliasing effects and a Krylov-space inversion method to enable high performance with high resolution. These techniques are implemented in the ARGO computer code. We reconstruct the field over a 500 Mpc cube with Mpc grid-resolution while accounting both for the angular and radial selection functions of the SDSS, and the shot noise giving an effective resolution of the order of $\sim$10 Mpc. In addition, we correct for the redshift distortions in the linear and nonlinear regimes in an approximate way. We show that the commonly used method of inverse weighting the galaxies by the corresponding selection function heads to excess noise in regions where the density of the observed galaxies is small. It is more accurate and conservative to adopt a Bayesian framework in which we model the galaxy selection/detection process to be Poisson-binomial. This results in heavier smoothing in regions of reduced sampling density. Our results show a complex cosmic web structure with huge void regions indicating that the recovered matter distribution is highly non-Gaussian. Filamentary structures are clearly visible on scales up to $\sim$20 Mpc. We also calculate the statistical distribution of density after smoothing the reconstruction with Gaussian kernels of different radii r_S and find good agreement with a log-normal distribution for $10\,\mathrm{Mpc} \lesssim r_S \lesssim 30\,\mathrm{Mpc}$.

Key words: large-scale structure of Universe – galaxies: clusters: general – methods: data analysis – methods: statistical

1 INTRODUCTION

Measuring the Large-Scale Structure (LSS) of the Universe has become a major task in cosmology in recent years. The relics of the seed fluctuations, originating from the inflationary phase of the early Universe, are mainly encoded in the linear regime of the LSS in which structure formation has not significantly degraded the primordial phase information. In particular there has recently been a focus on measuring the baryon acoustic signal imprinted in the galaxy distribution which has been suggested as a powerful standard ruler for our Universe (see for example Eisenstein 2005).

Upcoming and ongoing galaxy redshift surveys such as

DEEP2 or Baryon Oscillation Spectroscopic Survey (BOSS) will cover higher and higher redshifts (see for example Davis et al. 2005; Schlegel et al. 2007). They are designed to trace complex structures in the Universe and to study the environment of galaxies and their evolution.

We carry a reconstruction of the density field dealing with statistical and systematic errors of the galaxy distributions with the ARGO[1] computer code described in Kitaura & Enßlin (2008). ARGO is a high-performance implementation of a three-dimensional Wiener-filter, permitting treatments of an inhomogeneous and incomplete window function acting on the galaxy distribution. It exploits the power of fast Fourier transforms (FFTs) and

---

[*] E-mail: kitaura@sissa.it, kitaura@mpa-garching.mpg.de

[1] **A**lgorithm for the **R**econstruction of the **G**alaxy-traced **O**verdensity field

iterative Krylov-space based inversion schemes for the otherwise intractable data inversion step.

Reconstructions permit us to characterize the large-scale structure, helping to deepen our understanding of structure formation, to gain insight into the physical processes involved, to construct signal templates for the detection of weak physical effects. These can be used to study the cosmic microwave background and to reveal signals ranging from the Integrated Sachs-Wolfe effect (see for example Frommert et al. 2008), over the Sunyaev-Zel'dovich effect in the diffuse gas, to metal absorption lines. An interesting further application would be to constrain the bias between luminous and dark matter using reconstructions made by ARGO and correlating them with simulations and reconstructions of the matter distribution coming from other observables like weak lensing, Lyman alpha forest, etc. Topological studies could be made from the reconstructed data, leading to a geometrical characterization of the actual large-scale structure (see for example Sheth & Sahni 2005). It is also interesting to study how the physical properties of galaxies depend on their large-scale environment Li et al. (2006b); Lee & Lee (2008) and Lee & Li (2008). The reconstructed structures of a galaxy catalogue can be traced back in time with various methods, like those based on the Zel'dovich (1970) approximation (see for example Nusser & Dekel 1992). These early matter density fluctuations can be used as initial conditions for N-body simulations. The results of such a constrained simulation have a wide application in structure formation theory (see for example Mathis et al. 2002). A joint estimation of the matter field and its power-spectrum would also be a natural next step given the technology we develop below (for similar work in CMB-analysis see, for example, Wandelt et al. 2004; Jewell et al. 2004; Eriksen et al. 2007).

We present the first application of ARGO to observational data. In particular we have applied our method to recover the galaxy density field based on data from `Sample dr6fix` of the New York University Value Added Catalogue (NYU-VAGC), which was constructed from the Sloan Digital Sky Survey (SDSS; York et al. 2000) Data Release 6 (DR6; Adelman-McCarthy et al. 2008). This leads to the largest Wiener-reconstruction of the Large-Scale Structure made so far effectively requiring the inversion of a matrix with about $10^8 \times 10^8$ entries. The use of optimized iterative inversion schemes within an operator formalism (see Kitaura & Enßlin 2008) together with a careful treatment of aliasing effects (see Jasche et al. 2009) permits us to recover the overdensity field on Mpc scales (for previous Wiener reconstructions see Fisher et al. 1994; Hoffman 1994; Lahav et al. 1994; Lahav 1994; Zaroubi et al. 1995; Fisher et al. 1995; Webster et al. 1997; Zaroubi et al. 1999; Schmoldt et al. 1999; Erdoğdu et al. 2004, 2006). Note, that alternative density reconstruction techniques like Voronoi and Delaunay tesselations (see e.g. Icke & van de Weygaert 1991; Ebeling & Wiedenmann 1993; Zaninetti 1995; Bernardeau & van de Weygaert 1996; Doroshkevich et al. 1997; Meurs & Wilkinson 1999; Kim et al. 2000; Schaap & van de Weygaert 2000; van de Weygaert & Schaap 2001; Ramella et al. 2001; Panko & Flin 2004; Zaninetti 2006) are tuned to optimally represent the density field from a geometrical point of view, but are not explicit in the statistical assumptions made on the galaxy or matter distribution, which is an important aspect of our analysis here.

We investigate in detail the statistical problem of finding an expression for a noise covariance which includes the survey angular and radial selection functions. The expression we find assumes a binomial model for the galaxy selection/detection process.

We show that including our proposed noise covariance matrix in the Wiener-filter leads to a more conservative reconstruction of matter structures than using the inverse weighting scheme. We also compare the linear WF expression which is derived from a least squares approach and the non-linear WF which uses a signal dependent noise covariance (see appendix A in Kitaura & Enßlin 2008). The latter shows to be even more conservative than the linear WF since it strongly suppresses the cells with higher number counts.

Due to the fine mesh of the reconstruction (~ 1 Mpc) a treatment of the redshift distortion in the linear and non-linear regime is required. We choose a redshift distortion deconvolution method, as presented by Erdoğdu et al. (2004), which aims to correct in both regimes. This treatment only corrects the power and neglects any phase information. For this reason, the effective resolution of the reconstruction is lower than the resolution of the grid (~ 10 Mpc).

Our paper is structured as follows. We start by describing the input galaxy sample of the Sloan Digital Sky Survey (SDSS) Data Release 6 (DR6) in section 2. Then we present the methodology used to perform an estimation of the matter field (section 3). In detail, the galaxy distribution is first transformed into the comoving frame (section 3.1.1) and then assigned to a grid using our newly developed *supersampling* method (described in Jasche et al. 2009) to correct for aliasing effects, ensuring a correct spectral representation of the galaxy distribution even up to the highest modes contained in the grid (section 3.1.2). Completeness on the sky and radial galaxy selection function are then translated into a three dimensional mask, which will be part of the response operator used in the filtering step (section 3.1.3). Then, an observed galaxy overdensity field is calculated which fulfils the statistical requirements we want to impose on the matter field (section 3.2). Taking the observed galaxy field as the data vector we finally apply a Wiener-filtering step with the ARGO computer code (section 3.3.1) followed by a deconvolution step, effectively correcting for the redshift distortion (section 3.3.2). Here, we distinguish between a linear WF expression which is derived from a least squares approach and a non-linear WF which uses a signal-dependent noise covariance. Both WF formulations are tested with mock data and quantitatively compared to a simple procedure in which the galaxies are inverse weighted with the completeness, then gridded and finally smoothed to give a matter field estimate.

We present a reconstruction of the density field for the DR6 main sample in section 5. First, we analyze the survey sky mask (section 5.1). Results for the Sloan Great Wall are then presented in detail. Some other prominent structures, for example, the Coma, the Leo, and the Hercules clusters are also discussed (section 5.2) together with the detection of a large void region (section 5.3). The proper implementation of the filter enables us to deal with complex masks which include unobserved regions. We demonstrate the improved detection of overdensity regions close to edges of the mask and the prediction of structures in gaps, as demonstrated by comparing with data from the Data Release 7 (DR7) where those gaps are filled (section 5.4). In section 5.5 we analyze the statistical distribution of the density field and find good agreement with a log-normal distribution for smoothing radii in a Gaussian filter in the range $10\,\mathrm{Mpc} \lesssim r_\mathrm{S} \lesssim 30\,\mathrm{Mpc}$. Finally, we make a summary of the work, and present our conclusions and future outlook.

2 INPUT GALAXY SAMPLE

In this study we use data from the sixth data release (DR6; Adelman-McCarthy et al. 2008) of the Sloan Digital Sky Survey (SDSS; York et al. 2000). The survey contains images of a quarter of the sky obtained using a drift-scan camera (Gunn et al. 1998) in the *u, g, r, i, z* bands (Fukugita et al. 1996; Smith et al. 2002; Ivezić et al. 2004), together with spectra of almost a million objects obtained with a fibre-fed double spectrograph (Gunn et al. 2006). Both instruments were mounted on a special-purpose 2.5 meter telescope (Gunn et al. 2006) at Apache Point Observatory. The imaging data are photometrically (Hogg et al. 2001; Tucker et al. 2006) and astrometrically (Pier et al. 2003) calibrated, and were used to select spectroscopic targets for the main galaxy sample (Strauss et al. 2002), the luminous red galaxy sample (Eisenstein et al. 2001), and the quasar sample (Richards et al. 2002). Spectroscopic fibres are assigned to the targets using an efficient tiling algorithm designed to optimize completeness (Blanton et al. 2003c). The details of the survey strategy can be found in York et al. (2000) and an overview of the data pipelines and products is provided in the Early Data Release paper (Stoughton et al. 2002). More details on the photometric pipeline can be found in Lupton et al. (2001) and on the spectroscopic pipeline in SubbaRao et al. (2002).

We take data from `Sample dr6fix` of the New York University Value Added Catalogue (NYU-VAGC). This is an update of the catalogue constructed by Blanton et al. (2005) and is based on the SDSS DR6 data and publicly available selection masks[2]. Starting from `Sample dr6fix`, we construct a magnitude-limited sample of galaxies with spectroscopically measured redshifts in the range $0.001 < z < 0.4$, r-band Petrosian apparent magnitudes $14.5 < m \leqslant 17.6$, and r-band absolute magnitudes $-23 < M_{0.1_r} < -17$. Here m is corrected for Galactic extinction, and the apparent magnitude limits are chosen in order to get a sample that is uniform and complete over the entire area of the survey. The absolute magnitude $M_{0.1_r}$ is corrected to its $z = 0.1$ value using the K-correction code of Blanton et al. (2003a) and the luminosity evolution model of Blanton et al. (2003b). We also restrict ourselves to galaxies located in the main area of the survey in the northern Galactic cap, excluding the three survey strips in the southern cap, i.e. we include galaxies with right ascension (α) and declination (δ) in the following ranges: $105^{\circ} < \alpha < 270^{\circ}$ and $-5^{\circ} < \delta < 70^{\circ}$. In addition, we considered only galaxies which are inside a comoving cube of side 500 Mpc (with equal side lengths: $Lx \times Ly \times Lz$), as we describe below. These restrictions result in a final sample of 255,818 galaxies.

In order to correct for incompleteness in our spectroscopic sample, we need to have complete knowledge of its selection effects. A detailed account of the observational selection effects accompanies the NYU-VAGC release. These include two parts: a mask on the sky and a radial selection function along the line-of-sight. The mask shows which areas of the sky have been targeted, and which have not, either because they are outside the survey boundary, because they contain a bright confusing source, or because observing conditions were too poor to obtain all the required data. The effective area of the survey on the sky defined by this mask is 5314 square degrees for the sample we use here. It is divided into a large number of smaller subareas, called *polygons*, for each of which the NYU-VAGC lists a spectroscopic completeness.

This is defined as the fraction of the photometrically defined target galaxies in the polygon for which usable spectra were obtained. The average completeness over our sample galaxies is 0.86. The radial selection function gives the fraction of galaxies in the absolute magnitude range being considered ($-23 < M_{0.1_r} < -17$ in our case) that are within the apparent magnitude range of the sample ($14.5 < m \leqslant 17.6$ in our case) at a given redshift.

In certain cases we also work with a sample of galaxies drawn from SDSS data release 7 (DR7; Abazajian et al. 2009) for which the galaxy positions, redshifts and fluxes are publicly available from the SDSS website[3] but the survey completeness as described above was not released at the moment this work started. With this sample we apply only a gridding scheme and a subsequent Gaussian smoothing, without accounting for any selection effects, in order to *qualitatively* check for overdense regions present in the gap in the SDSS DR6.

3 METHODOLOGY

In this section, we describe the main algorithms required to perform a Wiener-filter reconstruction of the matter field as described in Kitaura & Enßlin (2008) (see also the pioneering works Wiener 1949; Rybicki & Press 1992; Zaroubi et al. 1995). We start with the preparation of the data followed by a filtering step and a final deconvolution. Detailed descriptions of the methodology used for each step are described in the following subsections.

3.1 Preparation of the data

Reconstructing a signal like the matter density field from the observed galaxy sample requires a model which relates the underlying matter field to the galaxy distribution. This model will define the inverse problem, which can be solved with a reconstruction algorithm. In this subsection, we describe how to prepare the input data in such a way that it is consistent with the data model underlying the ARGO-code.

3.1.1 Transformation of the data into comoving coordinates

To apply a reconstruction algorithm which uses the correlation function in comoving space, we first have to transform the redshift distances into comoving distances for each galaxy by performing the integral[4]:

$$r \equiv \int \mathrm{d}z \frac{1}{cH(z)}, \qquad (1)$$

with $H(z)$ being the Hubble parameter given by:

$$H(z) = H_0 \sqrt{\Omega_{\mathrm{m}}(1+z)^3 + \Omega_{\mathrm{K}}(1+z)^2 + \Omega_{\Lambda}}, \qquad (2)$$

where we chose the concordance ΛCDM-cosmology with $\Omega_{\mathrm{m}} = 0.24$, $\Omega_{\mathrm{K}} = 0$ and $\Omega_{\Lambda} = 0.76$ (Spergel et al. 2007). In addition, we assumed a Hubble constant: $H_0 = h\mathrm{km/s/Mpc}$ with $h = 73$.

With this definition the three-dimensional galaxy positions (X,Y,Z) in comoving space are calculated as follows:

$$
\begin{aligned}
X &= r \cdot \cos(\delta) \cdot \cos(\alpha) \\
Y &= r \cdot \cos(\delta) \cdot \sin(\alpha) \\
Z &= r \cdot \sin(\delta).
\end{aligned}
\qquad (3)
$$

---

[3] http://www.sdss.org/dr7
[4] Not to be confused with the r-band.

3.1.2 Supersampling step

Now, we can sort the galaxies onto a grid with a *supersampling* scheme, which will permit us to apply a reconstruction scheme based on FFTs. The much lower computational costs of FFTs permits us to tackle much more ambitious matter reconstructions than have been attempted previously with Wiener-filtering techniques. The main difficulty in signal processing via FFT techniques arises from the need to represent a continuous signal which extends to infinity on a finite discrete grid. Various methods to approximate the real continuous signal by a discrete representation have been proposed in literature, e.g. Nearest Grid Point (NGP), Cloud In Cell (CIC) or Triangular Shaped Clouds (TSC) (see e.g Hockney & Eastwood 1981). However, all of these methods are only approximations to the ideal low-pass filter, and introduce discretisation artifacts such as aliasing. For a detailed discussion see e.g. Hockney & Eastwood (1981); Jing (2005); Cui et al. (2008); Jasche et al. (2009). In recent years a number of methods have been proposed to correct for these artifacts, especially for the purpose of power-spectrum estimation (Jing 2005; Cui et al. 2008). However, common methods to suppress these artifacts in the discretised signals, tend to be numerically expensive.

To circumvent this problem, Jasche et al. (2009) proposed a supersampling technique, which is able to provide discrete signal representations with strongly suppressed aliasing contributions at reasonable computational cost. This method relies on a two-step filtering process, where in the first step the signal is pre-filtered by sampling the signal via the TSC method to a grid with twice the target resolution. In our case we use a 1024^3 grid. In a second step the ideal discrete low-pass filter is applied to the pre-filtered signal, allowing us to sample the low-pass filtered field at the lower target resolution. In this fashion we obtain an aliasing free signal sampled at a target resolution of 512^3 cells (with equal number of cells in each axis: $Nx \times Ny \times Nz$).

Let us define the observed galaxy sample as a point source distribution $n_{\mathrm{p}}^{\mathrm{o}}(s)$ with coordinate s

$$n_{\mathrm{p}}^{\mathrm{o}}(s) \equiv \sum_{i=1}^{N_{\mathrm{c}}^{\mathrm{o}}} \delta_{\mathrm{D}}(s - s_i), \qquad (4)$$

with $N_{\mathrm{c}}^{\mathrm{o}}$ being the total observed galaxy number count and δ_{D} the Dirac-delta function. The process of putting the galaxies on a regular grid is equivalent to a convolution in real-space followed by a grid-point selection step according to Hockney & Eastwood (1981)

$$n^{\mathrm{o}}(s) \equiv \Pi\Big(\frac{s}{H}\Big) \int \mathrm{d}s' \, K_{\mathrm{S}}(s - s') n_{\mathrm{p}}^{\mathrm{o}}(s'), \qquad (5)$$

with $\Pi(r) = \sum_{n \in \mathbb{Z}} \delta_{\mathrm{D}}(s - n)$, H being the grid-spacing and K_{S} the supersampling kernel. We define the resulting field as the observed galaxy number density $n^{\mathrm{o}}(s)$. The observed galaxy number density is a function of the Cartesian position in comoving space, but includes redshift distortion. For this reason, we say that the distribution is in redshift-space denoted by the coordinate s.

3.1.3 Calculation of the three-dimensional mask: completeness on the sky and selection function

To define the data vector we need to model the three dimensional mask. We do this by processing the two-dimensional sky mask in several steps. First, the sky mask or completeness on the sky $w_{\mathrm{SKY}}(\alpha, \delta)$ is evaluated using the survey mask provided in `Sample dr6fix` of the NYU-VAGC (see Section 2) on an

equidistant $\alpha \times \delta$-grid with 165000×75000 cells having a resolution of $36''$ both in right ascension and declination (see panel (a) in Fig. 6). Then, we project the sky mask on a comoving Cartesian $X \times Y \times Z$-grid containing 512^3 cells.

This is done with the transformation given by Eqs. 3 taking projected values of the mask every 0.25 Mpc in the radial direction which are then assigned on the grid using the Nearest Grid Point (NGP) method and normalized by the number of mask counts at that grid cell. The analogous procedure is done with the radial completeness $w_r(z)$, i.e. the selection function which is available as a function of redshift.

Finally, we obtain the three dimensional mask $w(s)$ as a product of the projected two dimensional mask, i.e. the completeness on the sky $w_{\mathrm{SKY}}(\alpha, \delta)$ and the projected selection function $w_r(s)$ (see Fig. 1 and panels a in figures 6, 8, 9 and 10 in section 5). We define $w(s) \leqslant 1$.

3.2 Definition of the data model

Let us define the observed galaxy overdensity field as[5]:

$$\delta_{\mathrm{g}}^{\mathrm{o}}(s) \equiv \frac{n^{\mathrm{o}}(s)}{\overline{n}} - w(s), \qquad (6)$$

with $\overline{n}$ being the mean galaxy number density.

The mean galaxy number density on the grid $\overline{n}$ is defined by the quotient of the total number of observed galaxies $N_{\mathrm{c}}^{\mathrm{o}}$ and the observed volume V^{o}. Note, that this assumes that the observed volume is a **fair sample** of the Universe. We can then write:

$$\overline{n} \equiv \frac{N_{\mathrm{c}}^{\mathrm{o}}}{V^{\mathrm{o}}} \equiv \frac{\sum_{i=1}^{N_{\mathrm{cells}}} N_{\mathrm{ci}}^{\mathrm{o}}}{\int \mathrm{d}r \, w(r)}, \qquad (7)$$

with $N_{\mathrm{ci}}^{\mathrm{o}}$ being the number of observed galaxies at cell i: $N_{\mathrm{c}}^{\mathrm{o}} \equiv \sum_{i=1}^{N} N_{\mathrm{ci}}^{\mathrm{o}}$, N_{cells} being the total number of cells and the observed volume being defined by the integral: $V^{\mathrm{o}} \equiv \int \mathrm{d}r \, w(r)$. The relation between the expected galaxy number density in a small volume ΔV around position r $\rho_{\mathrm{g}}(r)$ and the mean galaxy number density in the whole volume under consideration V is given by:

$$\rho_{\mathrm{g}}(r) \equiv \overline{n}(1 + \delta_{\mathrm{g}}(r)), \qquad (8)$$

where $\delta_{\mathrm{g}}(r)$ is the galaxy overdensity field, which describes the spatial density distribution of galaxies. Here we assume that effects due to galaxy evolution are negligible in the observed region, and, especially, that the mean number density is redshift independent.

The observed quantity $\delta_{\mathrm{g}}^{\mathrm{o}}(s)$ defined in Eq. 6 has to be related to the signal, we seek to recover, via a data model. This relation is to be inverted by the reconstruction algorithm.

3.2.1 Physical model

In this section we describe the physical model which will enable us to apply linear reconstruction methods and obtain an estimate of the matter field valid on large-scales (>1 Mpc). Let us assume a continuous matter field $\delta_{\mathrm{m}}(r)$ in comoving space r as well as a continuous galaxy field δ_{g}. We model the actual galaxies as being Poisson distributed according to this field with an expectation density of $\overline{n}(1 + \delta_{\mathrm{g}}(r))$. In general, the relation between the galaxy overdensity field δ_{g} and the underlying matter field δ_{m} will be given by a non-local and nonlinear bias operator. However, the formalism

---

[5] Not to be confused with the declination δ.

we present here, without any further development, allows us to account only for a non-local linear translation-invariant bias operator $B(\boldsymbol{r} - \boldsymbol{r}')$ of the form:

$$\delta_{\mathrm{g}}(\boldsymbol{r}) \equiv \int \mathrm{d}\boldsymbol{r}'\, B(\boldsymbol{r} - \boldsymbol{r}')\delta_{\mathrm{m}}(\boldsymbol{r}'). \qquad (9)$$

Note, that this linear operator is known to fail at least at sub-Mpc scales. Several non-local biasing models are described in the literature, which are mainly used to correct for the shape of the power-spectrum on large-scales (Tegmark et al. 2004; Hamann et al. 2008). We will carry this general bias through the algebraic calculations. However, in this work we consider the galaxy field to be a fair sample of the matter field. Thus, we assume the special case of a linear constant bias equal to unity: $B(\boldsymbol{r}, \boldsymbol{r}') = \delta_{\mathrm{D}}(\boldsymbol{r} - \boldsymbol{r}')$, so that $\delta_{\mathrm{g}} = \delta_{\mathrm{m}}$. Nevertheless, any non-local bias scheme of the form of Eq. 9 can be adapted without the need to repeat the filtering. We show that one can easily deal with non-local bias models in a final deconvolution step (see Eq. 30). As a result, various posterior biasing assumptions can be applied based on this reconstruction to test different biasing models.

We will also assume the existence of a redshift distortion operator[6] $Z(\boldsymbol{s}, \boldsymbol{r})$, which transforms the density field from real-space into redshift-space. Note, that the redshift distortion operator cannot be a linear operator, since it depends on the matter field $\delta_{\mathrm{m}}(\boldsymbol{r})$. However, we will approximate it with a linear redshift distortion operator $Z(\boldsymbol{s}, \boldsymbol{r})$ here:

$$\delta_{\mathrm{g}}(\boldsymbol{s}) \equiv \int \mathrm{d}\boldsymbol{r}\, Z(\boldsymbol{s}, \boldsymbol{r})\delta_{\mathrm{g}}(\boldsymbol{r}), \qquad (10)$$

and postpone a matter field dependent treatment, sampling the peculiar velocity field as proposed in (Kitaura & Enßlin 2008), for later work.

Let us further assume an additive noise term resulting in a data model for the observed galaxy overdensity as:

$$\delta_{\mathrm{g}}^{\mathrm{o,th}}(\boldsymbol{s}) \equiv w(\boldsymbol{s})\int \mathrm{d}\boldsymbol{r}\, Z(\boldsymbol{s}, \boldsymbol{r})\int \mathrm{d}\boldsymbol{r}'\, B(\boldsymbol{r} - \boldsymbol{r}')\delta_{\mathrm{m}}(\boldsymbol{r}') + \epsilon(\boldsymbol{s}), \qquad (11)$$

with ϵ being the noise term. The corresponding vector representation of the data model can be approximated as:

$$\boldsymbol{\delta}_{\mathrm{g},s}^{\mathrm{o,th}} \equiv \mathbf{W}_s \mathbf{Z}_{s,r} \mathbf{B}_r \boldsymbol{\delta}_{\mathrm{m},r} + \boldsymbol{\epsilon}_s, \qquad (12)$$

with the subscripts r and s denoting real-space and redshift-space, respectively. The response operator can be defined by

$$\mathbf{R}_{s,r} \equiv \mathbf{W}_s \mathbf{Z}_{s,r} \mathbf{B}_r, \qquad (13)$$

with $\mathbf{W}_s$ being the three dimensional mask operator defined in continuous space by: $W(\boldsymbol{s}, \boldsymbol{s}') = w(\boldsymbol{s})\delta_{\mathrm{D}}(\boldsymbol{s} - \boldsymbol{s}')$, $\mathbf{Z}_{s,r}$ being the redshift distortion operator, and $\mathbf{B}_r$ being the bias operator. Now we need to specify a model for the noise term.

3.2.2 Statistical model

Assuming that the galaxy distribution is generated by an inhomogeneous Poissonian distribution, the number galaxy count N_{c} within a volume ΔV around position $\boldsymbol{r}$ is distributed as:

$$N_{\mathrm{c}}(\boldsymbol{r}) \sim P_{\mathrm{Pois}}(N_{\mathrm{c}}(\boldsymbol{r})|\lambda(\boldsymbol{r})). \qquad (14)$$

with

$$P_{\mathrm{Pois}}(N_{\mathrm{c}}(\boldsymbol{r})|\lambda(\boldsymbol{r})) = \frac{\lambda(\boldsymbol{r})^{N_{\mathrm{c}}(\boldsymbol{r})}}{N_{\mathrm{c}}(\boldsymbol{r})!}\exp(-\lambda(\boldsymbol{r})), \qquad (15)$$

where the expected number of galaxy counts is given by the Poissonian ensemble average: $\lambda(\boldsymbol{r}) = \langle N_{\mathrm{c}}(\boldsymbol{r})\rangle_{\mathrm{g}}$ and is directly related to the expected galaxy density ρ_{g} at that position: $\rho_{\mathrm{g}}(\boldsymbol{r}) \equiv \langle N_{\mathrm{c}}(\boldsymbol{r})\rangle_{\mathrm{g}}/\Delta V$. Here $\langle\, \}\rangle_{\mathrm{g}} \equiv \langle\{\, \}\rangle_{(N_{\mathrm{c}}|\lambda)} \equiv \sum_{N_{\mathrm{c}}=0}^{\infty} P_{\mathrm{Pois}}(N_{\mathrm{c}} \mid \lambda)\{\, \}$ denotes an ensemble average over the Poissonian distribution. We further model the observational selection of $N_{\mathrm{c}}^{\mathrm{o}}(\boldsymbol{r})$ galaxies out of the N_{c} present within the small volume ΔV to be a binomial selection with an acceptance rate $w(\boldsymbol{r})$. We then can write:

$$N_{\mathrm{c}}^{\mathrm{o}}(\boldsymbol{r}) \sim P_{\mathrm{Bin}}(N_{\mathrm{c}}^{\mathrm{o}}(\boldsymbol{r})|N_{\mathrm{c}}(\boldsymbol{r}), w(\boldsymbol{r})), \qquad (16)$$

with

$$P_{\mathrm{Bin}}(N_{\mathrm{c}}^{\mathrm{o}}(\boldsymbol{r}) \mid N_{\mathrm{c}}(\boldsymbol{r}), w(\boldsymbol{r}))$$
$$= \binom{N_{\mathrm{c}}(\boldsymbol{r})}{N_{\mathrm{c}}^{\mathrm{o}}(\boldsymbol{r})}(w(\boldsymbol{r}))^{N_{\mathrm{c}}^{\mathrm{o}}(\boldsymbol{r})}(1 - w(\boldsymbol{r}))^{(N_{\mathrm{c}}(\boldsymbol{r}) - N_{\mathrm{c}}^{\mathrm{o}}(\boldsymbol{r}))}.$$

The expected mean observed number of galaxies in the volume ΔV is:

$$\langle N_{\mathrm{c}}^{\mathrm{o}}(\boldsymbol{r})\rangle_w = w(\boldsymbol{r})N_{\mathrm{c}}(\boldsymbol{r}), \qquad (17)$$

where $\langle\{\, \}\rangle_w \equiv \langle\{\, \}\rangle_{(N_{\mathrm{c}}^{\mathrm{o}}|N_{\mathrm{c}}, w)} \equiv \sum_{N_{\mathrm{c}}^{\mathrm{o}}=0}^{\infty} P_{\mathrm{Bin}}(N_{\mathrm{c}}^{\mathrm{o}} \mid N_{\mathrm{c}}, w)\{\, \}$ represents the ensemble average over the binomial distribution with a selection probability w. Consequently, one can model the observed number of galaxies, as a single Poissonian process:

$$N_{\mathrm{c}}^{\mathrm{o}}(\boldsymbol{r}) \sim P_{\mathrm{Pois}}(N_{\mathrm{c}}^{\mathrm{o}}(\boldsymbol{r})|\lambda^{\mathrm{o}}(\boldsymbol{r})), \qquad (18)$$

with mean

$$\lambda^{\mathrm{o}}(\boldsymbol{r}) \equiv w(\boldsymbol{r})\lambda(\boldsymbol{r}) = w(\boldsymbol{r})\langle N_{\mathrm{c}}(\boldsymbol{r})\rangle_{\mathrm{g}} = \langle\langle N_{\mathrm{c}}^{\mathrm{o}}(\boldsymbol{r})\rangle_{\mathrm{g}}\rangle_w. \qquad (19)$$

Note, that the Poissonian and the binomial distributions commute with each other.

3.2.3 Noise covariance and data autocorrelation matrix

Let us define the noise covariance matrix, according to the assumptions made in the previous section, as the shot noise resulting from an **inhomogeneous Poisson distribution** for the galaxy distribution $n(\boldsymbol{s})$, and a **binomial distribution** for describing the observation process which reduces the fraction of observed galaxies following the selection function. We then obtain an expression for the noise covariance[7]:

$$N^{\mathrm{SD}}(\boldsymbol{s}_1, \boldsymbol{s}_2) \equiv \langle \epsilon(\boldsymbol{s}_1)\epsilon(\boldsymbol{s}_2)\rangle_{(\epsilon|\delta_{\mathrm{m}}, \boldsymbol{p}_\epsilon)} \equiv \langle\langle \epsilon(\boldsymbol{s}_1)\epsilon(\boldsymbol{s}_2)\rangle_{\mathrm{g}}\rangle_w$$
$$\equiv \frac{1}{\bar{n}^2}(\langle\langle n^{\mathrm{o}}(\boldsymbol{s}_1)n^{\mathrm{o}}(\boldsymbol{s}_2)\rangle\rangle_{\mathrm{g}}\rangle_w - \langle\langle n^{\mathrm{o}}(\boldsymbol{s}_1)\rangle_{\mathrm{g}}\rangle_w\langle\langle n^{\mathrm{o}}(\boldsymbol{s}_2)\rangle_{\mathrm{g}}\rangle_w)$$
$$= \frac{1}{\bar{n}^2}\langle\langle n^{\mathrm{o}}(\boldsymbol{s}_1)\rangle_{\mathrm{g}}\rangle_w\delta_{\mathrm{D}}(\boldsymbol{s}_1 - \boldsymbol{s}_2)$$
$$= \frac{1}{\bar{n}^2}w(\boldsymbol{s}_1)\langle n(\boldsymbol{s}_1)\rangle_{\mathrm{g}}\delta_{\mathrm{D}}(\boldsymbol{s}_1 - \boldsymbol{s}_2), \qquad (20)$$

where we have used the properties of the variance and mean of these distribution functions and have added the superscript SD to denote that this covariance matrix is signal-dependent (see section 2.5.3 and appendix A in Kitaura & Enßlin 2008). Note, that this noise covariance is defined as the ensemble average of the correlation matrix of the noise over all possible noise realizations denoted by the subscript $(\epsilon \mid \boldsymbol{\delta}_{\mathrm{m}}, p_\epsilon)$ with $\boldsymbol{p}_\epsilon$ being a set of parameters which determine the noise. Here, we have neglected the cell to cell correlation introduced by the gridding scheme we have used (TSP) as the first step in our *supersampling* scheme.

---

[6] Not to be confused with the Z-axis in our Cartesian grid.

[7] Not to be confused with the galaxy number counts N_{c}.

Having defined the data model, together with the noise model, we can calculate the expected data autocorrelation matrix, which is defined as the ensemble average over all possible galaxy realizations and density realizations (cosmic variance) leading to the following expression:

$$\langle\langle\langle\delta_{\rm g}^{\rm o,th}(\boldsymbol{s}_1)\delta_{\rm g}^{\rm o,th}(\boldsymbol{s}_2)\rangle_w\rangle_{\rm g}\rangle_{\rm m} \tag{21}$$
$$= \; w(\boldsymbol{s}_1)w(\boldsymbol{s}_2)\int {\rm d}\boldsymbol{r}_1\, Z(\boldsymbol{s}_1,\boldsymbol{r}_1)\int {\rm d}\boldsymbol{r}_2\, Z(\boldsymbol{s}_2,\boldsymbol{r}_2)$$
$$\times \int {\rm d}\boldsymbol{r}_1'\, B(\boldsymbol{r}_1-\boldsymbol{r}_1')\int {\rm d}\boldsymbol{r}_2'\, B(\boldsymbol{r}_2-\boldsymbol{r}_2')\,\langle\delta_{\rm m}(\boldsymbol{r}_1')\delta_{\rm m}(\boldsymbol{r}_2')\rangle_{\rm m}$$
$$+\langle N(\boldsymbol{s}_1,\boldsymbol{s}_2)\rangle_{\rm m},$$

with $\langle\{\,\}\rangle_{\rm m}\equiv\langle\{\,\}\rangle_{(\delta_{\rm m}|\boldsymbol{p}_{\rm m})}\equiv\int {\rm d}\delta_{\rm m}P(\delta_{\rm m}\mid\boldsymbol{p}_{\rm m})$ being the ensemble average over all possible matter density realizations with some prior distribution $P(\delta_{\rm m}\mid\boldsymbol{p}_{\rm m})$ with $\boldsymbol{p}_{\rm m}$ being a set of parameters which determine the matter field, say the cosmological parameters. Note, that this equation is only valid in the approximation where the bias and the redshift distortion operators are linear.

The noise term is the in Eq. 21 has the following form:

$$N^{\rm LSQ}(\boldsymbol{s}_1,\boldsymbol{s}_2)\equiv\langle N^{\rm SD}(\boldsymbol{s}_1,\boldsymbol{s}_2)\rangle_{\rm m}\equiv\langle\epsilon(\boldsymbol{s}_1)\epsilon(\boldsymbol{s}_2)\rangle_{(\delta_{\rm m},\epsilon|\boldsymbol{p})}$$
$$\equiv\frac{1}{\overline{n}}w(\boldsymbol{s}_1)\delta_{\rm D}(\boldsymbol{s}_1-\boldsymbol{s}_2), \tag{22}$$

since $\langle\langle n(\boldsymbol{r})\rangle_{\rm g}\rangle_{\rm m}=\langle\overline{n}\,(1+\delta_{\rm g}(\boldsymbol{r}'))\rangle_{\rm m}=\overline{n}$, assuming again, that the observed volume is a **fair sample** of the Universe. The noise covariance has been denoted with the superscript LSQ because it corresponds to the expression which is obtained by performing the LSQ approach to derive the WF, i.e. minimising the ensemble average of the squared difference between the real underlying density field $\delta_{\rm m}$ and the LSQ estimator $\delta_{\rm m}^{\rm LSQ}$ over all possible signal $\delta_{\rm m}$ and noise ϵ realizations: $\langle(\delta_{\rm m}-\delta_{\rm m}^{\rm LSQ})^2\rangle_{(\delta_{\rm m},\epsilon|\boldsymbol{p})}$ with $\boldsymbol{p}$ being the joint set of parameters: $\boldsymbol{p}\equiv\{\boldsymbol{p}_{\rm m},\boldsymbol{p}_\epsilon\}$ (for a derivation see appendix B in Kitaura & Enßlin 2008). We have also assumed that the cross terms between the noise and the signal are negligible: $\langle\delta_{\rm m}\epsilon^\dagger\rangle_{\rm m}=0$. This should be further analyzed in future work. Higher order correlations between noise and signal in fact exist, and can be exploited using schemes like the Poissonian scheme proposed in Kitaura & Enßlin (2008). Note, however that we consider a signal-dependent noise for the WF Eq. 20 which requires a model for the expected observed galaxy number density $\langle\langle n^{\rm o}(\boldsymbol{s}_1)\rangle_{\rm g}\rangle_w$ (for differences in the derivation see Kitaura & Enßlin 2008). We restrict ourselves to the LSQ noise covariance model $N^{\rm LSQ}$ given by Eq. 22 in our application to the SDSS data (section 5). Note, that the LSQ representation of the Wiener-filter is a linear operator in contrast to the alternative formulation which depends on the signal and thus is a nonlinear filter. We explore methods to deal with the signal-dependent noise formulation with mock galaxy catalogues and compare the results to the LSQ version of the Wiener-filter (see section 4).

Note, that by construction the data autocorrelation matrices for the observed galaxy overdensity field and the theoretical overdensity field are identical given the noise model in Eq. 20:

$$\langle\langle\langle\delta_{\rm g}^{\rm o,th}(\boldsymbol{s}_1)\delta_{\rm g}^{\rm o,th}(\boldsymbol{s}_2)\rangle_{\rm g}\rangle_w\rangle_{\rm m}=\langle\langle\langle\delta_{\rm g}^{\rm o}(\boldsymbol{s}_1)\delta_{\rm g}^{\rm o}(\boldsymbol{s}_2)\rangle_{\rm g}\rangle_w\rangle_{\rm m}. \tag{23}$$

3.3 Reconstruction algorithm

In this section we propose a two step reconstruction process: first a Wiener-filter step and second a deconvolution step.

3.3.1 Wiener-filtering

First, we recover the galaxy field in redshift-space ($\delta_{{\rm g},s}$) applying the Wiener-filter. The version of the Wiener-filter we use can be derived as follows. Let us approximate the posterior distribution assuming a Gaussian prior and a Gaussian likelihood:

$$P(\delta_{{\rm g},s}\mid\delta_{{\rm g},s}^{\rm o},\boldsymbol{p})\propto \tag{24}$$
$$\exp\left(-\frac{1}{2}\left[\delta_{{\rm g},s}^{\;\dagger}\mathbf{S}_{{\rm g},s}^{\;-1}\delta_{{\rm g},s}+(\delta_{{\rm g},s}^{\rm o}-\mathbf{W}_s\delta_{{\rm g},s})^\dagger\mathbf{N}_s^{\;-1}(\delta_{{\rm g},s}^{\rm o}-\mathbf{W}_s\delta_{{\rm g},s})\right]\right),$$

with the signal autocorrelation matrix $\mathbf{S}_{{\rm g},s}\equiv\langle\delta_{{\rm g},s}(\delta_{{\rm g},s})^\dagger\rangle$ being the inverse Fourier transform of the assumed model galaxy power-spectrum in redshift-space: $\hat{\hat{S}}_{{\rm g},s}(\boldsymbol{k},\boldsymbol{k}')\equiv(2\pi)^3 P_{\rm g}^s(\boldsymbol{k}')\delta_{\rm D}(\boldsymbol{k}-\boldsymbol{k}')$ and the hats denoting the Fourier transform of the signal autocorrelation matrix. Note, that the posterior distribution depends also on a set of parameters $\boldsymbol{p}$ which determine the power-spectrum $P_{\rm g}^s(\boldsymbol{k})$. The log-posterior distribution is then given by:

$$\log P(\delta_{{\rm g},s}\mid\delta_{{\rm g},s}^{\rm o},\boldsymbol{p})\propto \tag{25}$$
$$\delta_{{\rm g},s}^{\;\dagger}\mathbf{S}_{{\rm g},s}^{\;-1}\delta_{{\rm g},s}+(\delta_{{\rm g},s}^{\rm o}-\mathbf{W}_s\delta_{{\rm g},s})^\dagger\mathbf{N}_s^{\;-1}(\delta_{{\rm g},s}^{\rm o}-\mathbf{W}_s\delta_{{\rm g},s})$$
$$=\delta_{{\rm g},s}^{\;\dagger}\mathbf{S}_{{\rm g},s}^{\;-1}\delta_{{\rm g},s}+\delta_{{\rm g},s}^{\;\dagger}\mathbf{W}_s^\dagger\mathbf{N}_s^{\;-1}\mathbf{W}_s\delta_{{\rm g},s}-\delta_{{\rm g},s}^{\;\dagger}\mathbf{W}_s^\dagger\mathbf{N}_s^{\;-1}\delta_{{\rm g},s}^{\rm o}$$
$$-\delta_{{\rm g},s}^{\rm o}{}^\dagger\mathbf{N}_s^{\;-1}\mathbf{W}_s\delta_{{\rm g},s}+\delta_{{\rm g},s}^{\rm o}{}^\dagger\mathbf{N}_s^{\;-1}\delta_{{\rm g},s}^{\rm o}.$$

The first two terms can be combined to one term: $\delta_{{\rm g},s}^{\;\dagger}(\boldsymbol{\sigma}_{\rm WF}^2)^{-1}\delta_{{\rm g},s}$, using the Wiener-variance: $\boldsymbol{\sigma}_{\rm WF}^2\equiv(\mathbf{S}^{-1}+\mathbf{W}_s^\dagger\mathbf{N}_s^{-1}\mathbf{W}_s)^{-1}$. To find the mean of the posterior distribution we seek an expression for the log-posterior of the form:

$$\log P(\delta_{{\rm g},s}\mid\delta_{{\rm g},s}^{\rm o},\boldsymbol{p})\propto(\delta_{{\rm g},s}-\langle\delta_{{\rm g},s}\rangle_{\rm WF})^\dagger(\boldsymbol{\sigma}_{\rm WF}^2)^{-1}(\delta_{{\rm g},s}-\langle\delta_{{\rm g},s}\rangle_{\rm WF}), \tag{26}$$

with $\langle\delta_{{\rm g},s}\rangle_{\rm WF}=\mathbf{F}_{\rm WF}\delta_{{\rm g},s}^{\rm o}$ being the mean after applying the Wiener-filter $F_{\rm WF}$ to the data. Now the third and the fourth term of eq. (26) can be identified with the terms in eq. (26) as:

$$-\delta_{{\rm g},s}^{\;\dagger}\mathbf{W}_s^\dagger\mathbf{N}_s^{\;-1}\delta_{{\rm g},s}^{\rm o}=-\delta_{{\rm g},s}^{\;\dagger}(\boldsymbol{\sigma}_{\rm WF}^2)^{-1}\mathbf{F}_{\rm WF}\delta_{{\rm g},s}^{\rm o}, \tag{27}$$

and

$$-\delta_{{\rm g},s}^{\rm o}{}^\dagger\mathbf{N}_s^{\;-1}\mathbf{W}_s\delta_{{\rm g},s}=-\delta_{{\rm g},s}^{\rm o}{}^\dagger\mathbf{F}_{\rm WF}^\dagger(\boldsymbol{\sigma}_{\rm WF}^2)^{-1}\delta_{{\rm g},s}, \tag{28}$$

respectively. The remaining term depends only on the data and is thus factorized in the posterior distribution function as part of the evidence. From both eq. (27) and eq. (28) we conclude that the Wiener-filter has the form

$$\mathbf{F}_{\rm WF}=\boldsymbol{\sigma}_{\rm WF}^2\mathbf{W}_s^\dagger\mathbf{N}_s^{\;-1}=(\mathbf{S}^{-1}+\mathbf{W}_s^\dagger\mathbf{N}_s^{-1}\mathbf{W}_s)^{-1}\mathbf{W}_s^\dagger\mathbf{N}_s^{\;-1}. \tag{29}$$

The mean $\langle\delta_{{\rm g},s}\rangle_{\rm WF}$ of the posterior distribution defined by Eq. 24 can be obtained by:

$$\langle\delta_{{\rm g},s}\rangle_{\rm WF}=\left(\mathbf{S}_{{\rm g},s}^{-1}+\mathbf{W}_s^\dagger\mathbf{N}_s^{-1}\mathbf{W}_s\right)^{-1}\mathbf{W}_s^\dagger\mathbf{N}_s^{-1}\delta_{{\rm g},s}^{\rm o}. \tag{30}$$

We favor this signal-space representation[8] of the Wiener-filter with respect to the equivalent and more frequently used data-space representation in LSS reconstructions: $\langle\delta_{{\rm g},s}\rangle_{\rm WF}=\mathbf{S}_{{\rm g},s}\mathbf{W}_s^\dagger\left(\mathbf{W}_s\mathbf{S}_{{\rm g},s}\mathbf{W}_s^\dagger+\mathbf{N}\right)^{-1}\delta_{{\rm g},s}^{\rm o}$ (see for example Zaroubi et al. 1995), because it avoids instabilities which otherwise arise in our rapid algorithm for evaluating the filter.

Let us distinguish between the linear LSQ and the nonlinear signal-dependent noise formulation of the Wiener-filter. The first takes the matter field averaged noise, covariance Eq. 22 $N=$

---

[8] We use here the terminology introduced in Kitaura & Enßlin (2008).

N^{LSQ} and is used below when analyzing the SDSS data (see section 5). In the case of a signal-dependent noise: $N = N^{\mathrm{SD}}$ one needs an estimate of the expected observed galaxy number density $w(s)\lambda(s) \equiv \langle\langle n^{\mathrm{o}}(s)\rangle_{\mathrm{g}}\rangle_w$ (see Eq. 20 and section 3.4.2). Such an approach was done by Erdoğdu et al. (2004).

3.3.2 Deconvolution step

In the second reconstruction step, we deconvolve the galaxy field $\langle\boldsymbol{\delta}_{\mathrm{g},s}\rangle_{\mathrm{WF}}$ from the assumed redshift distortion and galaxy bias operators, obtaining an estimate for the underlying matter field in real-space:

$$\langle\boldsymbol{\delta}_{\mathrm{m},r}\rangle_{\mathrm{WF}} = \mathbf{B}_r^{-1}\mathbf{Z}_{r,s}^{-1}\langle\boldsymbol{\delta}_{\mathrm{g},s}\rangle_{\mathrm{WF}}. \qquad (31)$$

In this approximation, we can easily transform the reconstructed galaxy field into the matter field by just performing a final deconvolution with some scale-dependent bias of the form: $\hat{\tilde{B}}(\boldsymbol{k},\boldsymbol{k}') \equiv b(\boldsymbol{k})\delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}')$. As already mentioned above, our result should not be restricted to a single arbitrary chosen bias model. We therefore choose to recover the galaxy field by assuming a bias equal to unity from which matter reconstructions for all possible linear (and invertible) bias schemes can easily be constructed via Eq. 31. Note, that an alternative representation of the Wiener-filter which regularizes the bias and the redshift distortion operator when they are not be invertible, consists of including them in the response operator (Eq. 13) when calculating the Wiener-filter, leading to: $\langle\boldsymbol{\delta}_{\mathrm{m},r}\rangle_{\mathrm{WF}} = \left(\mathbf{S}_{\mathrm{m},r}^{-1} + \mathbf{R}_{r,s}^{\dagger}\mathbf{N}_s^{-1}\mathbf{R}_{s,r}\right)^{-1}\mathbf{R}_{r,s}^{\dagger}\mathbf{N}_s^{-1}\boldsymbol{\delta}_{\mathrm{g},s}^{\mathrm{o}}$.

3.3.3 Redshift distortion operator

Following Erdoğdu et al. (2004) we define the power-spectrum in redshift-space as the product of the power-spectrum in real-space and an effective redshift distortion factor given by the angle-averaged Kaiser factor[9] $K(\boldsymbol{k},\mu)$ times the damping Lorentzian factor $D(\boldsymbol{k},\mu)$:

$$P_{\mathrm{m}}^s(\boldsymbol{k}) \equiv \langle K(\boldsymbol{k},\mu)D(\boldsymbol{k},\mu)\rangle_{\mu} P_{\mathrm{m}}^r(\boldsymbol{k}), \qquad (32)$$

with $\mu = \boldsymbol{k}\cdot\boldsymbol{r}/(|\boldsymbol{k}||\boldsymbol{r}|)$. The Kaiser factor is given by (see Kaiser 1987):

$$K(\boldsymbol{k},\mu) \equiv (1+\beta\mu^2)^2, \qquad (33)$$

with β being the redshift distortion parameter which can be approximated by: $\beta \simeq \Omega_{\mathrm{m}}^{0.6}$ assuming a constant bias equal to unity and neglecting dark energy dependences (see Lahav et al. 1991). The Lorentzian damping factor is based on an exponential distribution in real-space for the pairwise peculiar velocity field and is given by:

$$D(\boldsymbol{k},\mu) \equiv \frac{1}{1+(k^2\sigma_v^2\mu^2)/2}, \qquad (34)$$

with $k \equiv |\boldsymbol{k}|$ and σ_v being the average dispersion velocity of the galaxies, which we assume to be $\sigma_v = 500$ km s$^{-1}H_0^{-1}$ (see for example Ballinger et al. 1996; Jing et al. 1998; Jing & Börner 2004; Li et al. 2006a).

We refer to Erdoğdu et al. (2004) for the angle-average expression of the product of the Kaiser factor and the damping factor. Consequently, we introduce the angular averaged redshift distortion operator defined as the square root of the factor in the previous

Eqs.:

$$\hat{\mathbf{Z}}(\boldsymbol{k},\boldsymbol{k}') \equiv \sqrt{\langle K(\boldsymbol{k}',\mu)D(\boldsymbol{k}',\mu)\rangle_{\mu}}\,\delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}'). \qquad (35)$$

By construction, this operator yields the correct power-spectrum modification for the translation from real- to redshift space[10].

Note, that this approximation is valid up to second order statistics, and gives only an effective solution to the redshift distortion due to the angular averaging. A proper solution would require a phase and direction dependent redshift distortion operator. If we assume that the galaxy bias is unity, we then can write the galaxy power-spectrum in redshift space as: $P_{\mathrm{g}}^s(\boldsymbol{k}') = \langle K(\boldsymbol{k},\mu)D(\boldsymbol{k},\mu)\rangle_{\mu} P_{\mathrm{m}}^r(\boldsymbol{k})$. Note, that this reduces the validity of our reconstruction to scales larger than the mesh resolution which is of about 1 Mpc to scales of about 10 Mpc. The power spectrum in real-space P_{m}^r is given by a nonlinear power-spectrum that also describes the effects of virialised structures with a halo term as given by Smith et al. (2003) at redshift $z = 0$. In addition to the cosmological parameters presented in section 3.1.1, we assume a spectral index $n_s = 1$. With each of the required operators defined, we can now apply our reconstruction algorithm as we demonstrate in the next section.

3.4 Signal-dependent noise formulation of the Wiener-filter

To apply the signal-dependent noise formulation of the Wiener-filter one needs to find estimators for the expected density field in the signal-dependent noise covariance (Eq. 20). We require either a good estimator for $\lambda^{\mathrm{o}}(\boldsymbol{r}) \equiv \langle\langle N_c^{\mathrm{o}}(\boldsymbol{r})\rangle_{\mathrm{g}}\rangle_w$ or for $\lambda(\boldsymbol{r}) \equiv \langle N_c(\boldsymbol{r})\rangle_{\mathrm{g}}$ since $\lambda^{\mathrm{o}}(\boldsymbol{r}) \equiv w(\boldsymbol{r})\lambda(\boldsymbol{r})$.

3.4.1 Flat prior assumption

The inverse weighting estimator used in previous works to estimate the noise covariance (see for example Erdoğdu et al. 2004) can be derived from the frequentist approach by assuming a flat prior for the overdensity distribution or equivalently infinite cosmic variance.

Let us start with Bayes theorem:

$$P(\lambda^{\mathrm{o}}|N_c^{\mathrm{o}}) = \frac{P(N_c^{\mathrm{o}}|\lambda^{\mathrm{o}})P(\lambda^{\mathrm{o}})}{P(N_c^{\mathrm{o}})}. \qquad (36)$$

The flat prior is defined as: $P(\lambda^{\mathrm{o}}) = c$, with c being a constant. The evidence is then given by

$$P(N_c^{\mathrm{o}}) = \int_0^{\infty} \mathrm{d}\lambda^{\mathrm{o}}\, P(N_c^{\mathrm{o}}|\lambda^{\mathrm{o}})c = c, \qquad (37)$$

since

$$\int_0^{\infty} \mathrm{d}\lambda^{\mathrm{o}}\, P(N_c^{\mathrm{o}}|\lambda^{\mathrm{o}}) = \int_0^{\infty} \mathrm{d}\lambda^{\mathrm{o}}\, \frac{(\lambda^{\mathrm{o}})^{N_c^{\mathrm{o}}}\mathrm{e}^{-\lambda^{\mathrm{o}}}}{N_c^{\mathrm{o}}!} = \frac{\Gamma(N_c^{\mathrm{o}}+1)}{N_c^{\mathrm{o}}!} = 1. \qquad (38)$$

Consequently, we obtain that the posterior distribution is equal to the likelihood

$$P(\lambda^{\mathrm{o}}|N_c^{\mathrm{o}}) = P(N_c^{\mathrm{o}}|\lambda^{\mathrm{o}}). \qquad (39)$$

---

[9] Not to be confused with the supersampling kernel K_{S}.

[10] Note, that we deviate here from Erdoğdu et al. (2004) in the order of the angular averaging and square root. An inspection of the power-spectrum corresponding to the reconstructions shows, however that only the prescription as implemented here leads to agreement with the nonlinear Smith et al. (2003) power-spectrum.
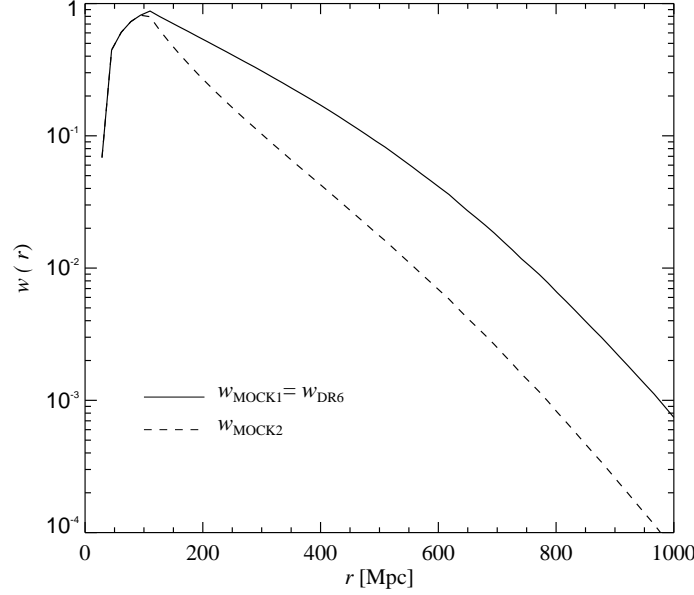
Figure 1. Radial selection functions used for the mock tests. Note, that the selection function used for the first mock test w_{MOCK1} is identical to the radial completeness of the DR6 catalogue w_{DR6}. The second selection function w_{MOCK2} is calculated by weighting $w_{\mathrm{DR6}}(r)$ with the factor $100 \mathrm{\ Mpc}/r$ for $r \geqslant 100$ Mpc.

The maximum likelihood estimator λ_{max} is obtained by looking at the extrema:

$$
\begin{aligned}
0 &= \frac{\partial P(\lambda_{\mathrm{max}}|N_c^o)}{\partial \lambda_{\mathrm{max}}} \\
&= (N_c^o(w\lambda_{\mathrm{max}})^{-1}w - w)\frac{(w\lambda_{\mathrm{max}})^{N_c^o}e^{-w\lambda_{\mathrm{max}}}}{N_c^o!} \\
&= N_c^o\lambda_{\mathrm{max}}^{-1} - w,
\end{aligned}
\tag{40}
$$

leading to:

$$
\lambda_{\mathrm{max}} = \frac{N_c^o}{w}.
\tag{41}
$$

Note, that the maximum estimator λ_{max} is not a valid estimator for the noise covariance matrix, since it can become zero at cells in which no galaxy count is present even if the cell belongs to the observed region. The mean estimator $\lambda_{\mathrm{mean}}^o$ can be found by performing the following integral:

$$
\begin{aligned}
\lambda_{\mathrm{mean}}^o &\equiv \int_0^{\infty} \mathrm{d}\lambda^o\, \lambda^o\, P(\lambda^o|N_c^o) \\
&= \int_0^{\infty} \mathrm{d}\lambda^o\, \frac{(\lambda^o)^{N_c^o+1}e^{-\lambda^o}}{(N_c^o+1)!}(N_c^o+1).
\end{aligned}
\tag{42}
$$

Thus, we have:

$$
\lambda_{\mathrm{mean}} \equiv \frac{\lambda_{\mathrm{mean}}^o}{w} = \frac{1}{w}(N_c^o+1).
\tag{43}
$$

The mean estimator λ_{mean} gives a regularized solution with respect to the maximum estimator λ_{max} overcoming the problem of having zero noise at cells with zero observed number counts. Both estimators however, rely on the flat prior assumption which can be dominated by the shot-noise for low completeness. This can be a problem when the reconstruction is performed on a fine mesh with extremely low completeness. For this reason, we test the SD Wiener-filter with an alternative scheme presented in the next section.

3.4.2 Statistically unbiased Jackknife-like scheme

The Jackknife-like scheme we present here and test in the next section produces subsamples from a galaxy distribution with selection function effects which are statistically unbiased with the underlying mean number density having a noise term with a structure function depending only on $\lambda(\boldsymbol{r})$. The first step of the scheme consists of generating a subsample using the binomial distribution given the observed number counts and the selection probability $\alpha/w(\boldsymbol{r})$ with a tunable parameter $\alpha < \min(w(\boldsymbol{r}))$:

$$
N_c'(\boldsymbol{r}) \sim \begin{cases} P_{\mathrm{Bin}}\left(N_c'(\boldsymbol{r}) \mid N_c^o(\boldsymbol{r}), \frac{\alpha}{w(\boldsymbol{r})}\right) \\ P_{\mathrm{Pois}}(N_c'(\boldsymbol{r}) \mid \alpha\lambda(\boldsymbol{r})). \end{cases}
$$

In the second step the subsample $N_c'(\boldsymbol{r})$ is inverse weighted with α:

$$
N_c''(\boldsymbol{r}) \equiv \frac{1}{\alpha}N_c'(\boldsymbol{r}).
\tag{44}
$$

One can notice, that the ensemble average over all possible α realizations leads to the mean number density $\lambda(\boldsymbol{r})$:

$$
\langle\langle N_c''(\boldsymbol{r})\rangle_{(N_c|\lambda)}\rangle_{\alpha} = \frac{1}{\alpha}\langle\langle N_c'(\boldsymbol{r})\rangle_{(N_c|\lambda)}\rangle_{\alpha} = \langle N_c(\boldsymbol{r})\rangle_{(N_c|\lambda)} = \lambda(\boldsymbol{r}).
\tag{45}
$$

Here, $\langle\{\}\rangle_{\alpha}$ is a binomial average with acceptance frequency α. The estimator for $\langle\lambda^o(\boldsymbol{r})\rangle_{JK} \equiv w(\boldsymbol{r})N_c''(\boldsymbol{r})$, with the subscript JK standing for the Jackknife estimator. We test the estimator proposed here to sample the noise covariance (see section 4).

4   QUALITY VALIDATION OF THE RADIAL SELECTION FUNCTION TREATMENT

In this section, we evaluate the quality of the reconstruction method under several incompleteness conditions. We restrict the study to a mesh of 128^3 cells for a cube with 500 Mpc side length and ignore
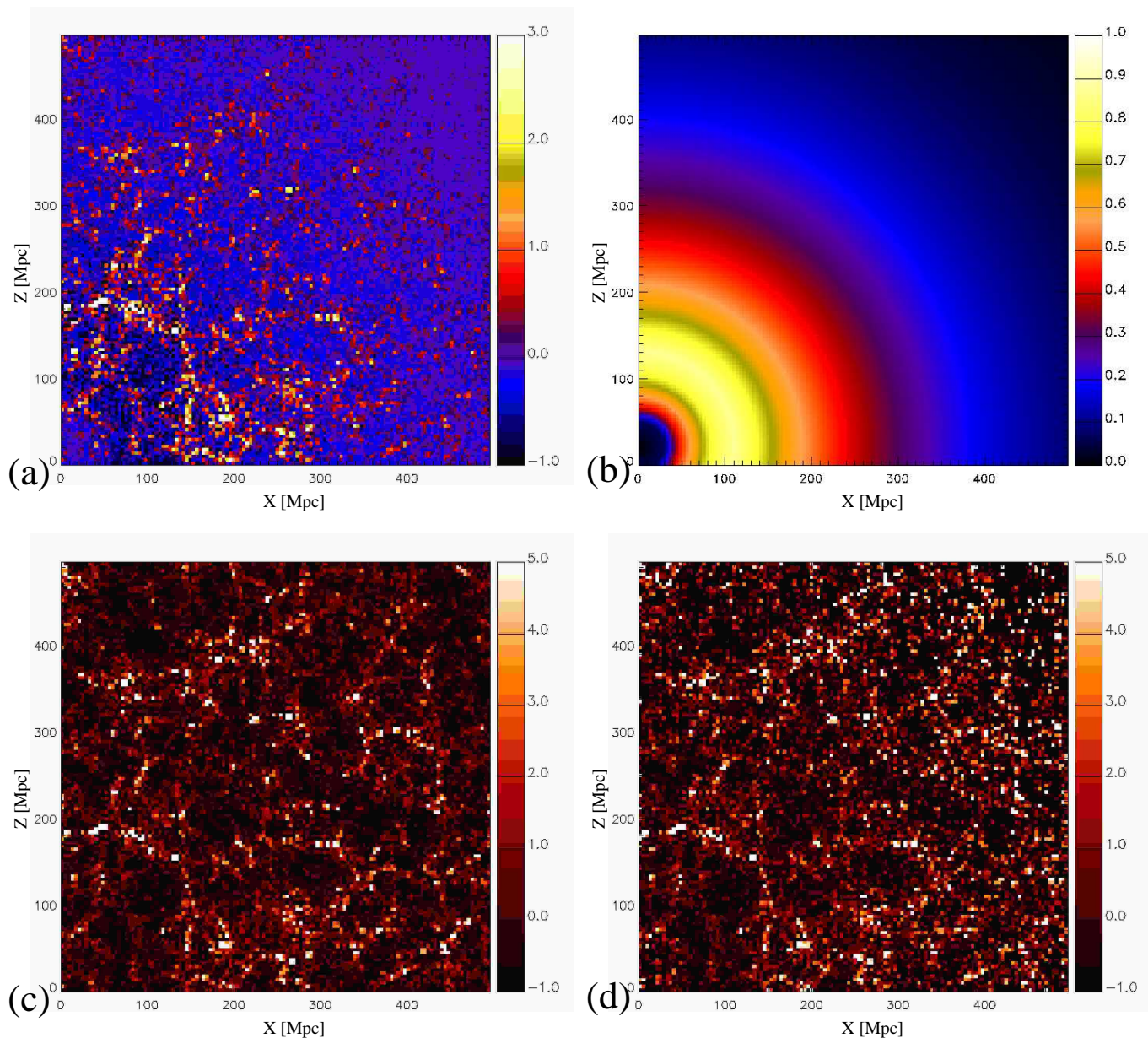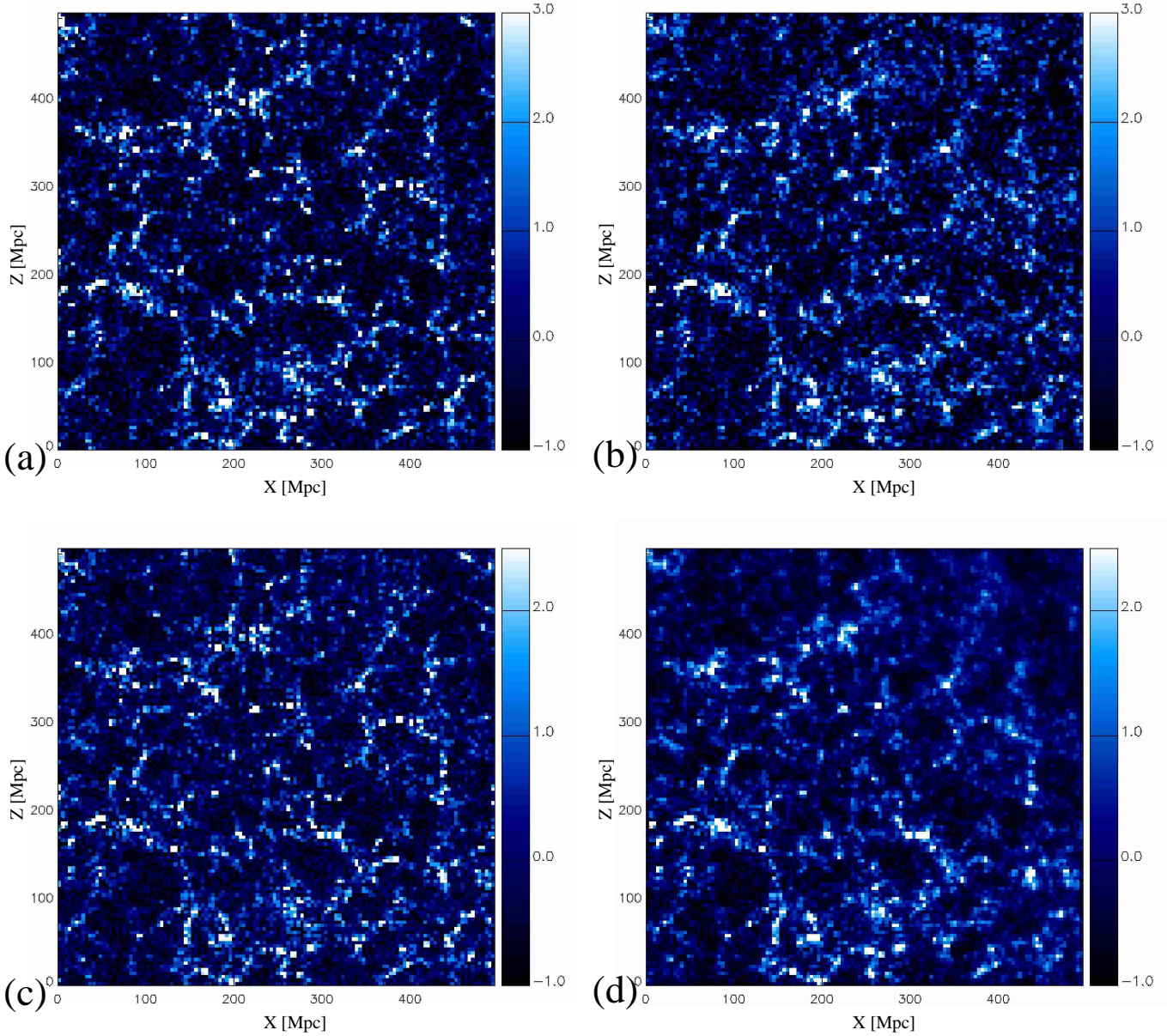
Figure 2. Mock test 1 using w_{MOCK1}. Input galaxy sample $\sim$ 20% of the complete galaxy sample. Slices around Y$\sim$ 270 Mpc through a 500 Mpc cube box with a 128^3 grid for different quantities without smoothing. Panel (a): observed mock galaxy overdensity field before correcting for the incompleteness. Panel (b): DR6 radial completeness corresponding to this test. Panel (c): underlying complete mock galaxy field. Panel (d): inverse weighting scheme applied to the sample represented in (a). Note, that panels (a), (c) and (d) were created taking the mean over 10 neighboring slices around the slice at Y$\sim$ 270 Mpc, corresponding to a thickness of 40 Mpc.

bias and redshift distortion effects. The necessity of performing a reconstruction step to make further studies of the large-scale structure is addressed. More simple schemes in which the galaxies are just gridded and the resulting field smoothed are shown to lead to significantly worse estimates of the matter field.

For this study, we consider a homogeneous subsample of 10^6 galaxies in a 500 Mpc cube box from the mock galaxy catalogue by De Lucia & Blaizot (2007) selected at random based on the Millennium Simulation (Springel et al. 2005). We define the 10^6 galaxy sample as our complete sample. Then, we generate two incomplete samples by radially selecting the galaxies according to two different radial completeness functions w_{MOCK1} and w_{MOCK2} (see

Fig. 1). This is done by drawing random uniform numbers between 0 and 1 for each mock galaxy and selecting the galaxies depending on whether the drawn number is above or below the value of the completeness at the corresponding distance to the observer. Note, that this ensures a perfect binomial observation process treating all the galaxies independent of their luminosity and thus avoiding the problem of galaxy biasing. The observer is defined in both cases at an equivalent position in the box to the real observer in the application to the observed DR6 data (section 5), namely at X=0 Mpc, Y=250 Mpc, and Z=20 Mpc. Note, that the arbitrary coordinates of the mock data range from 0 to 500 Mpc in each direction X, Y, and Z.

Figure 3. Mock test 1 using w_{MOCK1}. Input galaxy sample $\sim 20\%$ of the complete galaxy sample. Slices around Y~ 270 Mpc through a 500 Mpc cube box with a 128^3 grid for different quantities without smoothing. Panel (a): LSQ Wiener reconstruction to correct for the shot noise of the mock galaxy field taking the complete sample. Panel (b): LSQ Wiener reconstruction of the incomplete mock galaxy field taking into account the averaged shot noise and the radial selection function. Panel (c): mean over 200 Bayesian Wiener reconstructions to correct for the shot noise of the mock galaxy field taking the complete sample. Panel (d): mean over 200 SD Wiener reconstruction of the incomplete mock galaxy field taking into account shot noise and the radial selection function. Note, that all the panels were created taking the mean over 10 neighboring slices around the slice at Y~ 270 Mpc, i.e. over a slice of thickness 40 Mpc.

We consider the LSQ formulation of the Wiener-filter, which is a linear filter with a homogeneous noise term multiplied with a structure function given by the selection function Eq. 22, and the signal-dependent noise formulation, which is a nonlinear filter as it depends on the signal (see Eq. 20),and the inverse weighting scheme. In addition, to the Wiener-reconstruction methods, we define an inverse weighting scheme (IW) to estimate the underlying matter field as follows: first each galaxy is weighted with the inverse of the completeness at its location, then the galaxy sample is gridded according to the corresponding particle masses (we use our supersampling scheme to suppress aliasing), and finally the result-

ing field is convolved with different smoothing kernels. The first part of this scheme, leaving the smoothing for a later step, can be summarized by the following Eq.:

$$\langle n(\boldsymbol{r})\rangle_{\mathrm{IW}} \equiv \Pi\Big(\frac{\boldsymbol{r}}{H}\Big) \int \mathrm{d}\boldsymbol{r}'\, K_{\mathrm{S}}(\boldsymbol{r}-\boldsymbol{r}')\frac{1}{w(\boldsymbol{r}')} n_{\mathrm{P}}^{\mathrm{o}}(\boldsymbol{r}'), \quad (46)$$

where we have denoted the corresponding estimator by the angles: $\langle\{\,\}\rangle_{\mathrm{IW}}$. Note, that the completeness cannot be zero at a position in which a galaxy was observed. In order to make a quantitative comparison between the two Wiener-filtering methods and the inverse weighting method, a *true* underlying field δ^{true} needs to be defined. Since the inverse weighting scheme does not correct for the
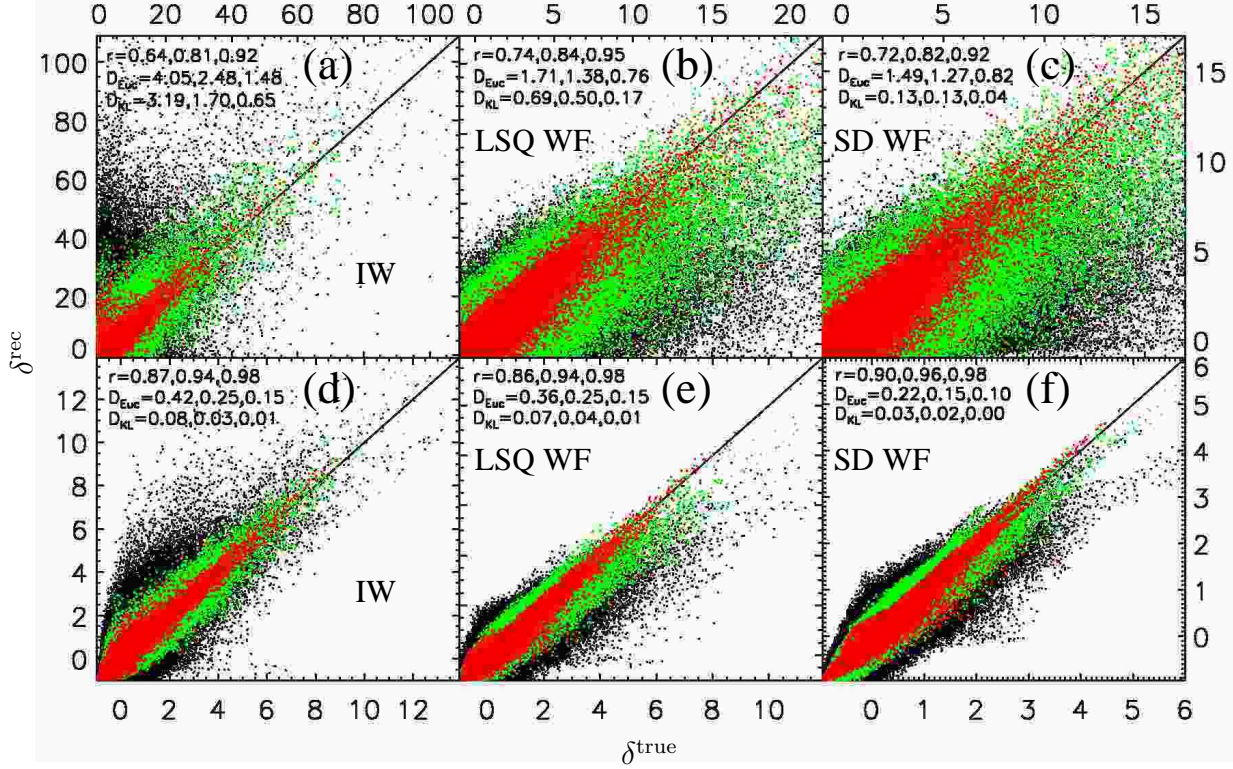
Figure 4. Statistical cell to cell correlation between the mock *true* density field δ^{true} and the reconstructed density field δ^{rec} at different scales for our first test case using w_{MOCK1}. Input galaxy sample $\sim 20\%$ of the complete galaxy sample. Also indicated: the statistical correlation coeficient r, the Euclidean distance D_{Euc} and the Kullback-Leibler distance D_{KL} first for all the sample (black dots), then for the sample in the radial comoving radius range between 200 and 400 Mpc (green dots), and finally in the range between 0 and 200 Mpc (red dots) away from the observer. The upper panels correspond to the comparison without smoothing and the lower panels after smoothing with a smoothing radius of $r_S = 5$ Mpc. Comparison between the complete mock galaxy field (in this case: δ^{true}) and the inverse weighting scheme applied to the incomplete sample (in this case: δ^{rec}) without smoothing (a) and after smoothing panel (d). Panel (b) and (e) represent the comparison between the average shot noise corrected complete mock galaxy field (in this case: δ^{true}) and the LSQ Wiener reconstruction of the incomplete sample (in this case: δ^{rec}) with the corresponding scale at bottom or top. Panel (c) and (f) represent the comparison between the local shot noise corrected complete mock galaxy field (in this case: δ^{true}) and the SD Wiener reconstruction of the incomplete sample using the Jackknife estimator (in this case: δ^{rec}) with the corresponding scale at bottom or top.

shot noise, we will compare with the complete mock galaxy sample (see panel (c) in Fig. 2) after smoothing on different scales. Note, that a consistent comparison for this case is difficult, since the shot noise varies with the different galaxy samples and with the distance to the observer. For the Wiener reconstruction case study we define the *true* underlying matter field δ^{true} as the resulting Wiener reconstruction taking the complete mock galaxy sample (see panel (e) in Fig. 2). Note that the *true* field thus also differs between our two Wiener filtering schemes. We will denote the reconstructed fields with each method as δ^{rec}.

4.1 Statistical correlation measures

To give a quantitative measurement of the quality of the reconstructions, we define the correlation coefficient r between the reconstructed and the *true* density field by[11]

$$\text{r}(\delta^{\text{rec}}, \delta^{\text{true}}) \equiv \frac{\sum_i^{N_{\text{cells}}} \delta_i^{\text{true}} \delta_i^{\text{rec}}}{\sqrt{\sum_i^{N_{\text{cells}}} (\delta_i^{\text{true}})^2} \sqrt{\sum_j^{N_{\text{cells}}} (\delta_j^{\text{rec}})^2}}. \quad (47)$$

The cell to cell plot of the reconstruction against the *true* density field is highly informative because the scatter in the alignment of the cells around the line of perfect correlation (45° slope) gives a qualitative goodness of the reconstruction. In general, the quality of the recovered density map is better represented by the Euclidean distance between the *true* and the reconstructed signal (see Kitaura & Enßlin 2008). The ensemble average of this quantity over all possible density realizations can also be regarded as an action or loss function that leads to the Wiener-filter through minimization (see Kitaura & Enßlin 2008). Here we introduce the Euclidean distance:

$$\text{D}_{\text{Euc}}(\delta^{\text{rec}}, \delta^{\text{true}}) \equiv \sqrt{\frac{1}{N_{\text{cells}}} \sum_i^{N_{\text{cells}}} (\delta_i^{\text{rec}} - \delta_i^{\text{true}}))^2}, \quad (48)$$

with $N_{\text{cells}} = 128^3$ for the mock tests). Let us, in addition, define the normalized Kullback-Leibler distance[12] (see

---

[11] Not to be confused with the comoving distance r.

[12] also called relative entropy in information theory
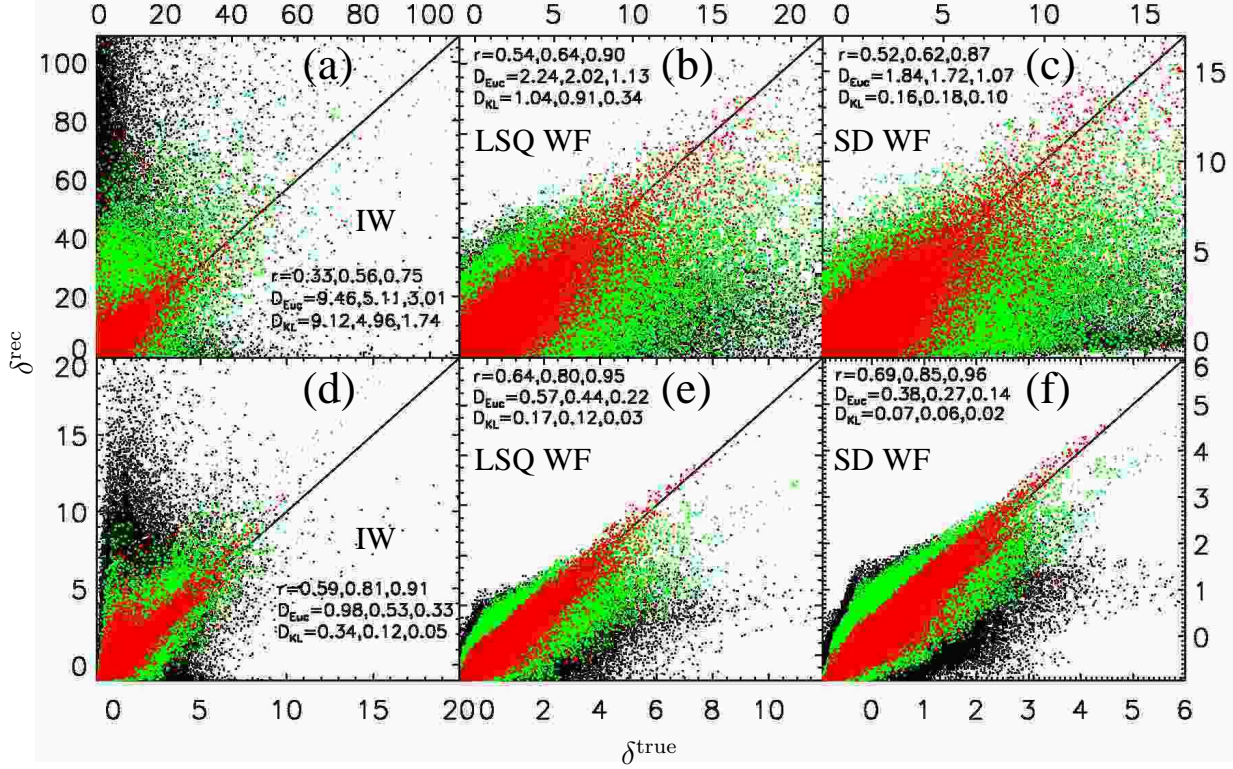
Figure 5. Same as Fig. 4, but using w_{MOCK2}.

Kullback & Leibler 1951) as

$$D_{\mathrm{KL}}(1+\delta^{\mathrm{rec}}, 1+\delta^{\mathrm{true}}) \equiv \frac{1}{N_{\mathrm{cells}}} \sum_{i}^{N_{\mathrm{cells}}} (1+\delta_i^{\mathrm{rec}}) \log\left(\frac{1+\delta_i^{\mathrm{rec}}}{1+\delta_i^{\mathrm{true}}}\right). \tag{49}$$

In our analysis we also compute smoothed versions of the density field convolving it with a Gaussian kernel given by:

$$G(\boldsymbol{r}, r_{\mathrm{S}}) \equiv \exp\left(\frac{|\boldsymbol{r}|^2}{2r_{\mathrm{S}}^2}\right), \tag{50}$$

with r_{S} being the smoothing radius.

4.2 First mock test

In the first mock test we try to emulate the same completeness conditions as given in the observed DR6 sample. For that, we take the complete mock galaxy catalogue (10^6 galaxies) and select according to the DR6 radial selection function ($w_{\mathrm{MOCK1}} = w_{\mathrm{DR6}}$) a subsample leaving about 20% of the total number of galaxies (218020) (see Fig. 2). The DR6 radial selection function can be seen as the black line in Fig. 1. A section through the box showing the completeness can be also seen in panel (b) of Fig. 2. The observer can be identified as being at the center of the spherical shells with equal completeness. The resulting overdensity field after applying this selection function to the complete mock sample can be seen in panel (a) of Fig. 2. Note, that we show here the mock observed galaxy field setting $w = 1$ in Eq. 6 in order to clearly see the selection effects. In the following, the discrete galaxy field (including Poisson noise) is represented with red color and the noise corrected field is represented in blue color. We will define the complete mock galaxy field including Poisson noise (panel (c) in Fig. 2) as the *true* galaxy

density field for the inverse weighting scheme. The corresponding noise corrected fields using the LSQ WF (panel (a) in Fig. 3) and the SD WF (panel (b) in Fig. 3) are defined as the *true* galaxy density field for the Wiener reconstructions. The *true* dark matter field is approximately related to this via Eq. 9, however, here we want to exclude the complication of galaxy biasing.

Panel (d) in Fig. 2 shows the result after applying the inverse weighting scheme. Panels (b) and (d) of Fig. 3 show the respective reconstructions using the LSQ and the SD WF. One can clearly see the noisy reconstruction produced by the inverse weighting scheme for structures located at large distances to the observer in contrast to the smoother estimation made by the Wiener-filtering schemes. The SD WF was applied for the complete galaxy sample using our statistically unbiased Jackknife-like scheme with an α parameter of 10^{-3}. The means after 200 reconstructions are shown in panels (c) and (d) for the complete and the selected samples respectively. The corresponding statistical analysis can be seen in Fig. 4. The cell to cell correlation plots show the tendency of the inverse weighting scheme to overestimate the density while the opposite is true in a significantly more moderate way when applying the Wiener-filter. In the case without smoothing (a mesh of size ~ 3.9 Mpc) (panels (a) and (d) in Fig. 4)) the qualitative and quantitative difference between the methods is very large, showing significantly better correlation coefficient and lower Euclidean and Kullback-Leibler distances for the Wiener reconstructions than for the inverse weighting scheme. Only when the fields are smoothed with a Gaussian of radius $r_{\mathrm{S}} = 5$ Mpc does the difference between the matter field estimators drop. With this smoothing the statistical correlation coefficient are similar for the Wiener-filter and the inverse weighting scheme. However, the Euclidean and Kullback-Leibler distances

remain being lower for the Wiener-filter (WF) reconstructions (see Fig. 4).

4.3 Second mock test

For the second mock test results we modify the DR6 selection function to drop faster towards larger radii leaving less than 10% of the galaxies (87220) by weighting $w_{DR6}(r)$ with the factor 100 Mpc/r for $r \geqslant 100$ Mpc. The corresponding radial selection function (w_{MOCK2}) can be seen as the dashed line in Fig. 1. The dramatic difference from DR6 completeness can be seen. using LSQ and SD formulations respectively. The noisy reconstruction produced by the inverse weighting scheme for structures located at large distances to the observer is now even more visible than in the previous test. Cells far away from the observed are excessively weighted. The Wiener-filter in contrast gives a smoother and more conservative estimation in regions in which the data are more incomplete. However, it remains sharp in regions where the information content is high (see structures close to the observer).

The corresponding statistical analysis can be seen in Fig. 5. The tendency to overestimate the density of the inverse weighting scheme is now extreme. Smoothing helps to raise the correlation coefficient values and to decrease the Euclidean and Kullback-Leibler distances. They remain, however, clearly above the ones achieved with the Wiener-filter schemes.

5 MATTER FIELD RECONSTRUCTIONS OF THE SDSS DR6

This work presents the first application of the ARGO-code to observational data. This yields the matter field reconstruction of the SDSS DR6 in the main area of the survey which is located in the northern Galactic cap on a comoving cube of side 500 Mpc and 512^3 cells.

In this section we describe a few remarkable features in the reconstructed matter field, demonstrating the quality of the reconstruction and the scientific potential for future applications. First, we discuss the mask and the projected three dimensional reconstruction without smoothing and after smoothing with a Gaussian kernel with a smoothing radius of r_S =5 Mpc and r_S =10 Mpc as displayed in Fig. 6. We then describe the largest structures in the nearby Universe, in particular the Sloan and CfA2 Great Walls (see Gott et al. 2005; Geller & Huchra 1989). Later, we analyze void or cluster detections which can be made with this kind of work. Finally, we analyze the statistical distribution of matter.

5.1 Mask and completeness

The sky mask for the region is shown in panel (a) of Fig. 6. The high resolution (36″ in both α and δ) permits us to visualize the plates of the SDSS with the intersection of several plates leading to higher completeness. The mask is divided into three patches: one small beam at high declination and right ascension angles and two wide regions. All the patches together cover almost a quarter of the sky. Between the two wider regions there is a large gap and there are several additional smaller gaps inside the patches. Such a complex mask is an interesting problem for the ARGO-code. It allows us to test, whether it can properly handle unobserved regions with zero completeness. Slices of the three dimensional mask calculated as the product of the completeness on the sky and the selection function (see section 3.1.3) are presented in panel (a) of Figs. 8,

9, 10 and panel (e) of Fig. 9. In these plots one can see how the selection function leads to a decrease of the completeness in the radial direction. Note, that the observer is located at (0,0,0) in our Cartesian coordinate system. We can see in panel (a) of Fig. 8 that the completeness rapidly reaches its maximum at around 110 Mpc distance from the observer and decreases at larger radii to values below 10%. In the next section we show how remarkably homogeneous structures are recovered in our reconstruction, independent of the distance from the observer and despite the low completeness values at large distances. We confirmed with additional reconstructions with larger volumes the same behavior for boxes up to side lengths of around 750 Mpc. For even larger volumes of 1 Gpc size, not shown here, however, the main sample becomes too sparse and only the large-scale structures are recovered. Including the three dimensional completeness for the SDSS DR6 data (see section 3.1.3) in Eq. 7 we obtain a mean galaxy density of about 0.05.

5.2 Mapping the Sloan and the CfA2 Great Wall

The Sloan Great Wall is one of the largest structure known in our local Universe although it is not a gravitationally bound object (see Gott et al. 2005). It extends for about[13] 400 Mpc (for a detailed study see Deng et al. 2006) and is located around 300 Mpc distant from Earth. In Fig. 7 we represent different radial shells, picking out the structures of the Sloan Great Wall, which extends from about 140° to 210° (-150° in Fig. 6) in right ascension and extends within a few degrees around declination $\delta \approx 0°$. In these shells other complex structures can be observed at higher declinations, showing filaments, voids and clusters of galaxies. Moreover, the region which has not been observed, lying outside the mask (see panel (a) in Fig. 6) is predicted to be filled with structures by the reconstruction method according to our assumed correlation function (see section 3.3.3). The Sloan Great Wall can also be seen in Fig. 8 almost in its full extent. We can see, how ARGO recovers the matter field, balancing the structures with low signal to noise ratio against those with a higher signal, leading to a homogeneously distributed field, meaning that clusters close to and far from the observer are both well represented. Only where the signal to noise drops below unity, do structures tend to *blur*, as can be observed in the upper parts of the reconstruction shown in Fig. 9.

The CfA2 Great Wall is also one of the largest structure known in our local Universe and contains the Coma Cluster (Abell 1656) at its center (see Geller & Huchra 1989). We can clearly see the Coma Cluster in the projected reconstruction without smoothing, being the big spot at right ascension $\alpha \approx 195°$ (-165° in Fig. 6) and declination $\delta \approx 28°$ in panel (b) of Fig. 6, located at a distance of ~ 100 Mpc from the observer (see Thomsen et al. 1997; Carter et al. 2008). The CfA2 Great Wall cannot be seen in its full extent in Fig. 8 because it reaches higher declination angles than selected in the plot. However, it can be partially seen as an elongated matter structure at about 100 Mpc distance to the observer, i.e. at around -100 Mpc in the X-axis in Fig. 8. Large filamentary structures are present even after smoothing with a Gaussian kernel with a smoothing radius of r_S =10 Mpc (see panel (d) in Fig. 8). The second major cluster of the Coma super-cluster is the Leo Cluster (Abell 1367) at a distance ~ 94 Mpc ($z \approx 0.022$), with galactic coordinates $\alpha \approx 176°$ and $\delta \approx 20°$. It is weakly detected in

---

[13] Note, that the extension of the Sloan Great Wall is usually given in luminosity distance, which can be around 40 Mpc larger than in comoving distance as we represent it here.
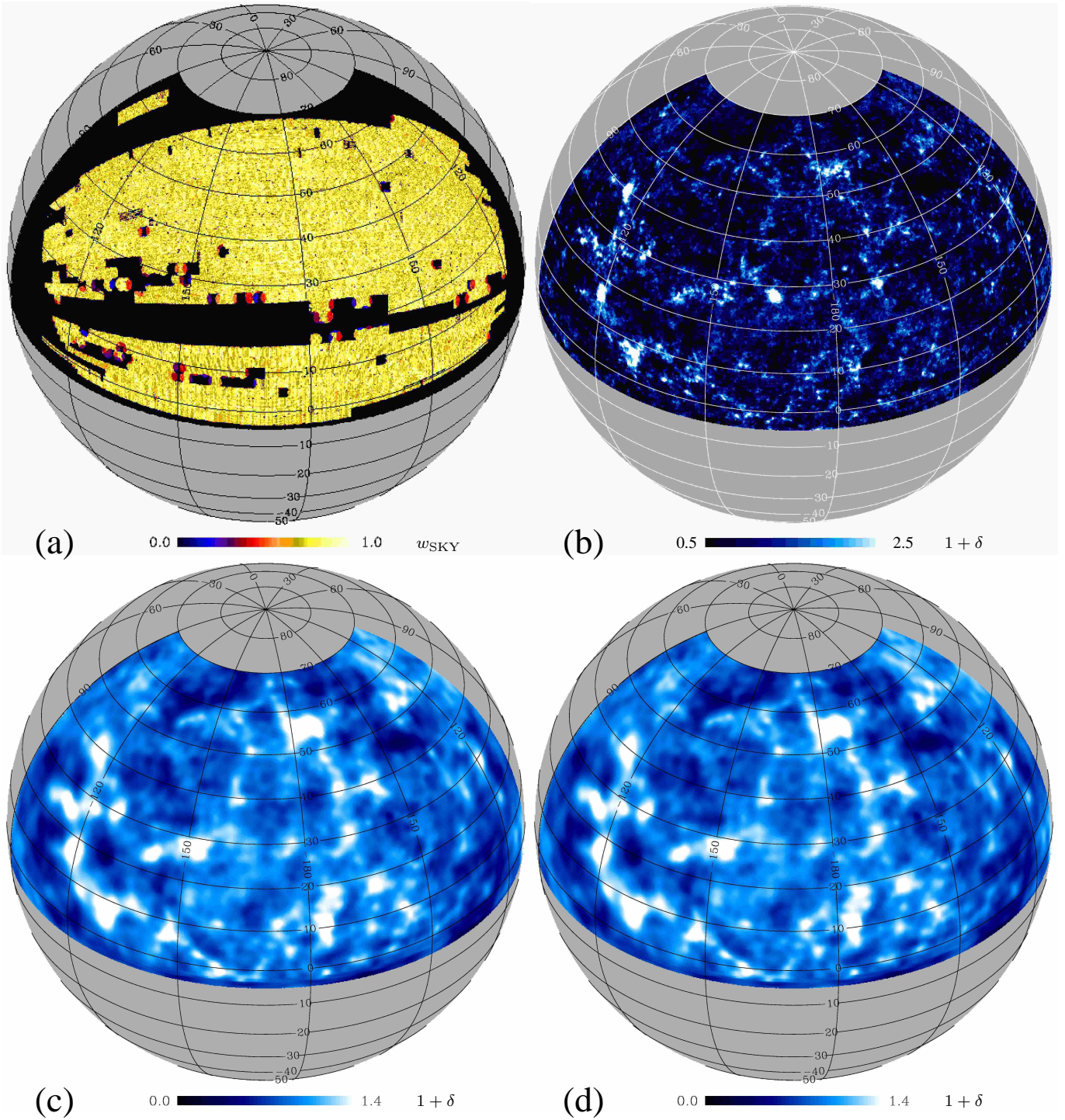
Figure 6. Panel (a): completeness of the observed patches on the sky. Shown are projections on the sky of the three dimensional matter field reconstruction, including the deconvolution with a redshift distortions operator and divided by the number of line-of sight grid-points used for the calculation to obtain a mean density field on the sky: without smoothing (panel (b)), after a convolution with a Gaussian kernel with a smoothing radius of $r_S = 5$ Mpc (panel (c)) and $r_S = 10$ Mpc (panel (d)). Note, that the longitude angles -90°, -120°, -150° and -180° correspond to 270°, 240°, 210° and 180° right ascension angles, respectively, with the positive angles being equal. For a general right ascension angle α the longitude is calculated as: α-360° for $\alpha \geqslant 180°$. The latitude angles are identical to the declination angles.

supercluster	cluster	Abell number	∼ right ascension α [degrees]	∼ declination δ [degrees]	∼ redshift
Coma	Coma	A1656	195° (-165°)	28°	0.0231
Coma	Leo	A1367	176° (176°)	20°	0.0220
Hercules		A2040	228° (-132°)	7°	0.0448
Hercules		A2052	229° (-131°)	7°	0.0338
Hercules		A2063	231° (-129°)	9°	0.0341
Hercules	Hercules	A2151	241° (-119°)	18°	0.0354
Hercules		A2147	241° (-119°)	16°	0.0338
Hercules		A2152	241° (-119°)	16°	0.0398
Hercules		A2148	241° (-119°)	25°	0.0418
Hercules		A2162	243° (-117°)	29°	0.0310
Hercules		A2197	247° (-113°)	41°	0.0296
Hercules		A2199	247° (-113°)	40°	0.0287

Table 1. Some of the most prominent clusters in the reconstruction with their corresponding right ascension and declination in degrees and redshift. Note, that the right ascension angle in Fig. 6 is indicated in parenthesis and can be calculated as: α-360° for $\alpha \geqslant 180°$.

our reconstruction as can be seen in panel (b) of Fig. 6, since it is partially located in the major gap of DR6 and should be therefore better detected with DR7.

The Hercules supercluster also belongs to the CfA2 Great Wall. Most of the clusters which belong this supercluster can be identified in the reconstructed area. Since the spatial range of these clusters is large, we have listed in table 1 the groups of clusters with their respective localisation in the sky which appear as especially prominent overdensity regions in the projected reconstruction (for references see Abell et al. 1989; Struble & Rood 1999). Note that close-by structures such as the Virgo Cluster, which is at a distance of only about 18 Mpc distance to us, cannot be detected in our reconstruction, because the lower limit of our sample is set at $z = 0.01$.

5.3 Detection of a great void region

The *scorpion*-like form of the matter distribution spanning the whole observed region in Fig. 9 (see mask in panel (a)) shows large connected filamentary structures with many clusters. Interestingly, an extremely large void is spanned in the region with -150 Mpc < Y < 30 Mpc and 70 Mpc < Z < 220 Mpc (see panels (a), (b) and (c) in Fig. 9). In order to evaluate the confidence of the detection one should check how deeply this region has been scanned by SDSS. By inspection of the three dimensional mask we confirm a fairly high completeness ranging from about 30% to about 65% (see panel (a) of Fig. 9). The extension in the X-axis is still unclear, since the gap in the mask grows in the void region to larger distances to the observer. ARGO predicts an extension of about -250 Mpc < X < -450 Mpc. From our results, we can tell that it is one of the largest voids in the reconstructed volume, having a diameter of about 150 Mpc. Conclusive results can only be obtained after investigating DR7, which fills the main gaps. Since, in this case, a proper treatment of the DR7 mask is required and this mask was not public at the time this project started, we postpone this study for

later work. The large overdensity region found in the unobserved region at about: -30 Mpc < Y < 30 Mpc and 370 Mpc < Z < 430 Mpc results from the correlation with a huge cluster region which extends in the range: -30 Mpc < Y < 30 Mpc and 350 Mpc < Z < 450 Mpc and which can be best seen at about X ∼ -170 Mpc (see panels (e) anf (f) in Fig. 9).

5.4 Cluster prediction

The signal-space representation of the Wiener-filter (see section 3.3.1) enables us to deal with unobserved regions, i.e. cells with zero completeness. Note, that for those cells the noise term vanishes in the Wiener-filter expression (Eq. 30). The filter can then be regarded as a convolution with the non-diagonal autocorrelation matrix of the underlying signal propagating the information from the windowed region into the unobserved cells. This gives a prediction for the Large-Scale Structure in these regions. Such an *extrapolation* can be clearly seen in panels (b), (c) and (d) of Fig. 6. These show the projected three dimensional reconstruction on the sky without smoothing and after a convolution with a Gaussian with a smoothing radius r_S of 5 and 10 Mpc, respectively. In these plots the gaps are hardly distinguishable, due to the signal prediction given by the Wiener-filter. We have chosen a slice, in which the propagation of the information through gaps can be analyzed. In panel (a) of Fig. 10 we can see the three-dimensional mask through our selected slice. The main gap crosses the entire box through the Y-axis and reaches about 50 Mpc width in the Z-axis. Several other smaller gaps are distributed in the slice. In the reconstruction in panel (b) we can see how the main gap is partially filled with some *diffuse* overdensity structures which are produced precisely as described above. Panel (c) shows the same reconstruction smoothed with a Gaussian kernel with a smoothing radius of $r_S = 5$ Mpc. Overplotted is the mask showing the regions in which it was observed. We identify seven clusters close to gaps extending into unobserved regions at a slice around -265 Mpc < X <
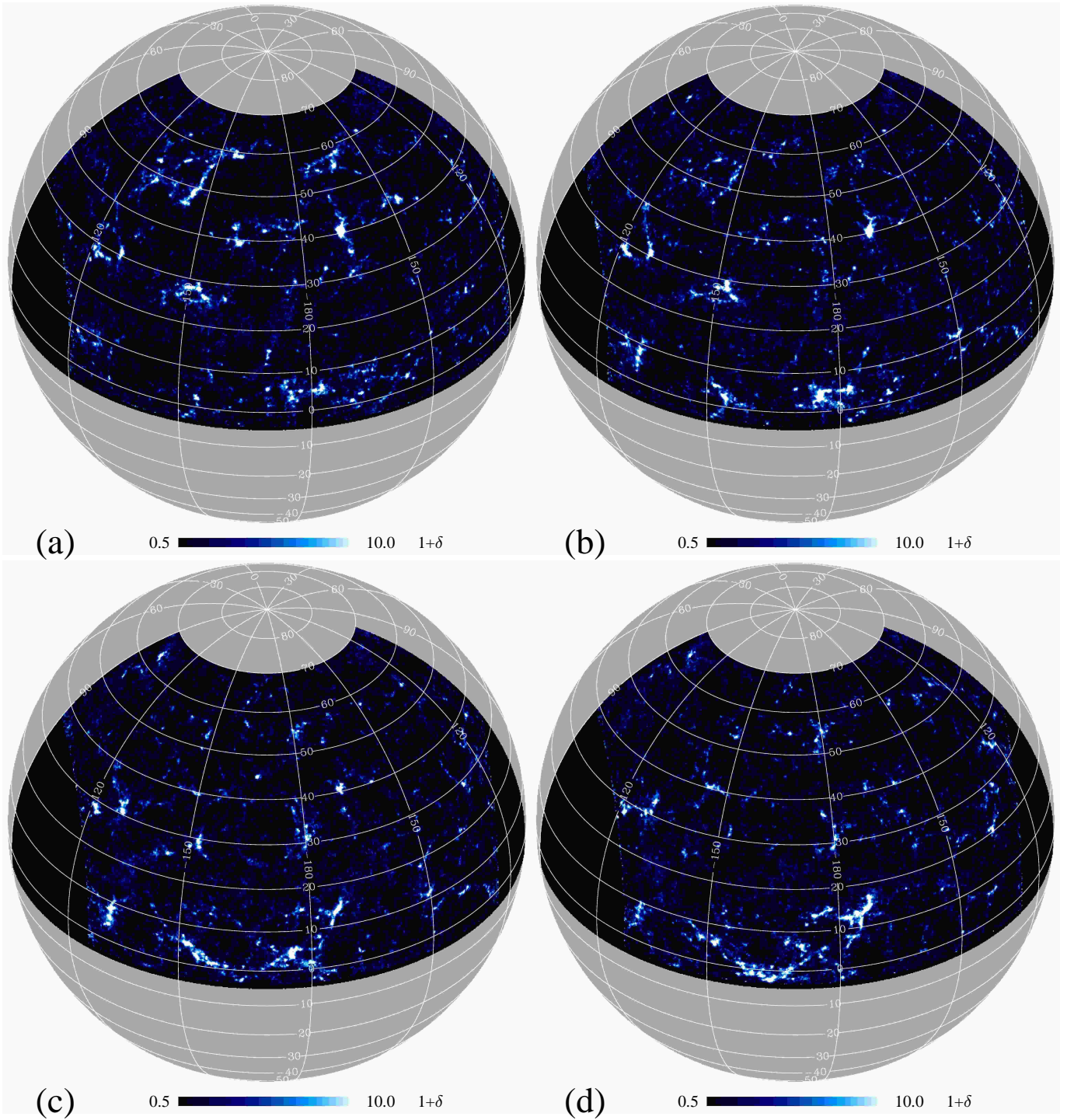
Figure 7. Different radial slices around the Sloan Great Wall. Shown are projections of the three dimensional matter field reconstruction on the sky considering only cells with a comoving distance between 290 Mpc and 310 Mpc (panel (a)), 300 Mpc and 320 Mpc (panel (b)), 310 Mpc and 330 Mpc (panel (c)), and 320 Mpc and 340 Mpc (panel (d)). Note, that the longitude angles -90°, -120°, -150° and -180° correspond to 270°, 240°, 210° and 180° right ascension angles, respectively, with the positive angles being equal. For a general right ascension angle α the longitude is calculated as: α-360° for $\alpha \geqslant 180°$. The latitude angles are identical to the declination angles.
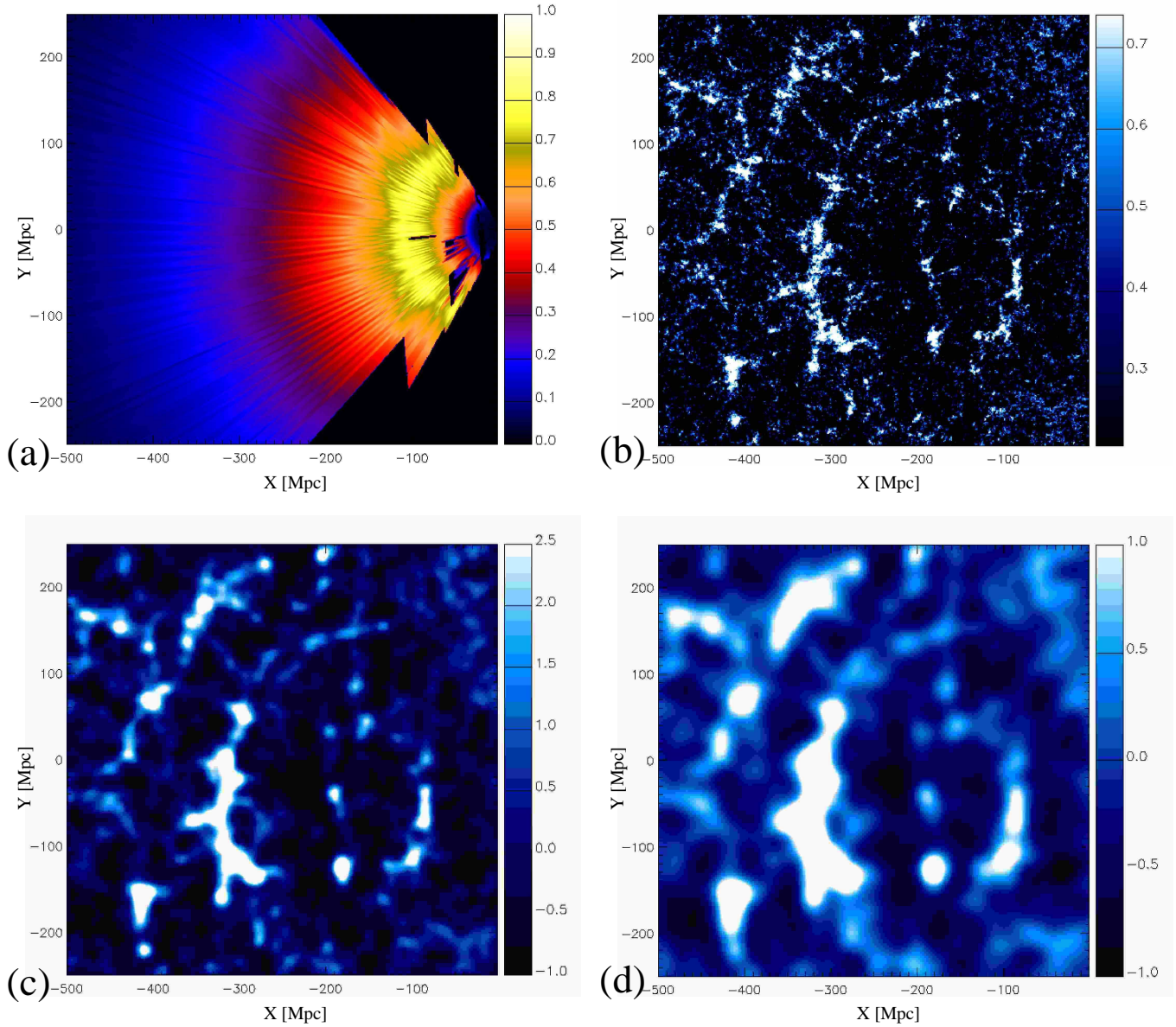
Figure 8. Slices around the Sloan and the CfA2 Great Wall. Panel (a): slice through the three dimensional mask multiplied with the selection function at $\sim$7 Mpc in the Z-axis. Panels (b), (c), and (d) show slices through the reconstruction after taking the mean over 20 neighboring slices around the slice at $\sim$7 Mpc in the Z-axis, without smoothing, convolved with a Gaussian kernel with a smoothing radius of r_S =5 Mpc and r_S =10 Mpc, respectively. Note, that panel (b) represents $\log(1 + \delta)$, whereas panels (c), and (d) show δ.

-245 Mpc (see clusters c_1-c_7 in Tab. 2). In addition, there are some weaker detections (see clusters c_8-c_{10} in Tab. 2). The gap which cluster c_1 extends into, and the largest gap, are the ones in which more information propagation occurs. There is an especially interesting region in the main gap around -140 Mpc < Y < 30 Mpc in which the algorithm predicts a high chance to find overdense structures. The rest of the gaps remains with low density values, since no prominent structures are in their vicinity. We investigate the public DR7 archive (see Section 2) to check for overdense regions in the gap. Note, that without a full angular and radial selection function treatment a quantitative comparison is not possible. We restrict our study by gridding the galaxy sample with NGP, ignoring mask or selection function effects, and convolving it with a Gaussian kernel with a smoothing radius of r_S =10 Mpc (see panel (d) in Fig. 10).

Though, faint features like the filaments lying at around -230 Mpc < Y < -130 Mpc cannot be recovered, stronger features like the clusters located at -100 Mpc < Y < 0 Mpc show that there is indeed an overdense region in the gap confirming our prediction based on DR6. In particular the extension of the clusters c_1 and c_2 are very well predicted by our algorithm. Cluster c_{10} is weakly predicted. The filament connecting clusters c_3 and c_{10} is predicted by ARGO, perhaps by chance, but the resemblance in the gap of the reconstruction to the real underlying distribution shows that use of the correlation function of the LSS allows for plausible predictions.
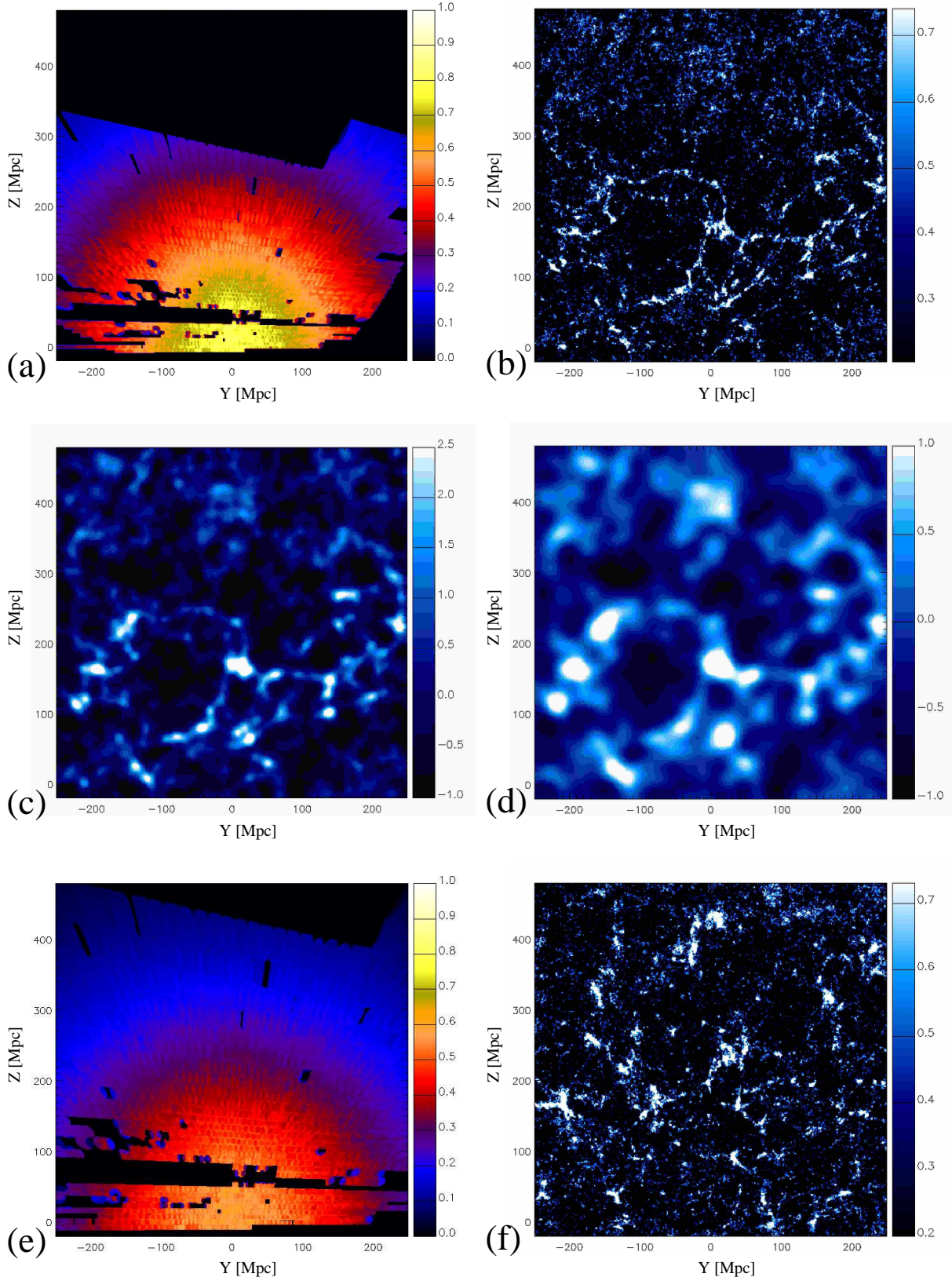
Figure 9. Panel (a): slice through the three dimensional mask multiplied with the selection function at ∼-109 Mpc in the X-axis. Panels (b), (c), and (d) show slices through the reconstruction after taking the mean over 20 neighboring slices around the slice at ∼-109 Mpc in the X-axis, without smoothing, convolved with a Gaussian kernel with a smoothing radius of r_S =5 Mpc and r_S =10 Mpc, respectively. Panel (e): slice through the three dimensional mask multiplied with the selection function at ∼-168 Mpc in the X-axis. Panel (f) shows a slice through the reconstruction after taking the mean over 20 neighboring slices around the slice at ∼-168 Mpc in the X-axis without smoothing. Note, that panels (b) and (f) represent $\log(1 + \delta)$, whereas panels (c), and (d) show δ.

c_1	-220 Mpc < Y < -200 Mpc	140 Mpc < Z < 180 Mpc
c_2	-140 Mpc < Y < -100 Mpc	120 Mpc < Z < 160 Mpc
c_3	10 Mpc < Y < 20 Mpc	120 Mpc < Z < 160 Mpc
c_4	10 Mpc < Y < 30 Mpc	70 Mpc < Z < 90 Mpc
c_5	60 Mpc < Y < 70 Mpc	70 Mpc < Z < 90 Mpc
c_6	150 Mpc < Y < 160 Mpc	60 Mpc < Z < 70 Mpc
c_7	220 Mpc < Y < 240 Mpc	90 Mpc < Z < 110 Mpc
c_8	-210 Mpc < Y < -200 Mpc	70 Mpc < Z < 90 Mpc
c_9	-110 Mpc < Y < -90 Mpc	70 Mpc < Z < 90 Mpc
c_{10}	-40 Mpc < Y < -60 Mpc	70 Mpc < Z < 90 Mpc

Table 2. Approximate positions of cluster candidates c_i (with i ranging from 1 to 10) at a slice around -265 Mpc < X < -245 Mpc in the reconstructed box which are located close to gaps (see Fig 10).

5.5 Statistics of the density field

From a physical point of view, one would expect a log-normal distribution of smoothed density for a certain range of smoothing scales, if one assumes an initial Gaussian velocity field and extrapolates the continuity equation for the matter flow into the nonlinear regime with linear velocity fluctuations (see Coles & Jones 1991). Since the log-normal field is not able to describe caustics, we expect this distribution to fail below a threshold smoothing scale. There should also be a transition at a certain scale between this quasi-linear regime and the linear regime where the matter field is still Gaussian distributed. Due to use of the Wiener-filter which considers only the correlation function to reconstruct the density field and the Gaussian smoothing, we expect the density field to be closely Gaussian distributed in the unobserved regions. Here, we analyse the statistical distribution of the density field by counting the number of cells at different densities with a density binning of 0.03 in $(1 + \delta_m)$ at different scales, defined by convolving the reconstruction with a Gaussian kernel with smoothing radii r_S of: 10, 20, and 30 Mpc. We performed the analysis for different radial shells in the Δr[14] ranges: $0 < r < 200$ Mpc, $200 < r < 400$ Mpc, $r > 400$ Mpc, and $0 < r < 600$ Mpc, separating observed ($w > 0$) and unobserved ($w = 0$) regions (see Figs. 11 and 12). Note, that due to shot noise, we are missing power in the filtered reconstruction on small scales. Moreover, the discrete Fourier representation of the signal implies negative densities (see Jasche et al. 2009). This obliges us to perform this statistical analysis on scales larger than the smallest grid scales. We can see this in the excess of low density cells for the dashed black curve (r_S =5 Mpc). In addition to that, we are also limited by the size of the box, having less information as we go to larger and larger scales. This effect can be appreciated in the stronger deviation from the log-normal fit around the peak for the green line (r_S =30 Mpc). For this reason, we restrict this analysis to the range of scales given above. The plots in Figs. 11 and 12 show how the distribution tends towards Gaussianity as we go to larger and larger scales.

We calculated the skewness and kurtosis to quantify the deviation from Gaussianity. Let us define here the statistical quantities required for our analysis. The number of cells contained in a shell of radial range Δr is given by the sum of the number counts in each density bin $f^B_{\Delta r, i}$:

$$N^{\Delta r}_{\text{cells}} \equiv \sum_i^{N^{\Delta r}_{\text{bins}}} f^B_{\Delta r, i}. \tag{51}$$

The mean overdensity in Δr which is very close to zero, is calculated as:

$$\overline{\delta^B_{\Delta r}} \equiv \frac{1}{N^{\Delta r}_{\text{cells}}} \sum_i^{N^{\Delta r}_{\text{bins}}} f^B_{\Delta r, i} \delta^B_{\Delta r, i}, \tag{52}$$

with the superscript ^{B} standing for bin. These two previously defined quantities permitted us to calculate the central n-moments μ_n of the distribution with:

$$\mu_n(\Delta r) \equiv \frac{1}{N^{\Delta r}_{\text{cells}}} \sum_i^{N^{\Delta r}_{\text{bins}}} f^B_{\Delta r, i} \left(\delta^B_{\Delta r, i} - \overline{\delta^B_{\Delta r}} \right)^n. \tag{53}$$

Note, that the variance is just the second moment: $\sigma^2 \equiv \mu_2$. Now, we can define the skewness[15]:

$$s \equiv \frac{\mu_3}{\sigma^3}, \tag{54}$$

and the kurtosis[16]:

$$k \equiv \frac{\mu_4}{\sigma^4} - 3. \tag{55}$$

Let us also introduce Pearson's skewness defined as the mean $\overline{\delta^B}$ minus the mode $\delta^B_{\max(f)}$ (overdensity bin with the maximum number of counts $\max(f)$) normalized by the square root of the variance:

$$s_P(\Delta r) \equiv \frac{\overline{\delta^B_{\Delta r}} - \delta^B_{\max(f^B(\Delta r))}}{\sigma(\Delta r)}. \tag{56}$$

The results are shown in Figs. 11 and 12 demonstrating large deviations from Gaussianity in the observed regions and negligible deviations for the unobserved regions. Since the Wiener filter uses only the first two moments of the matter distribution, we do not expect

---

[14] Note, that we considered the density at the center of the bins.

[15] Note, that for a Gaussian distribution: $s = 0$.
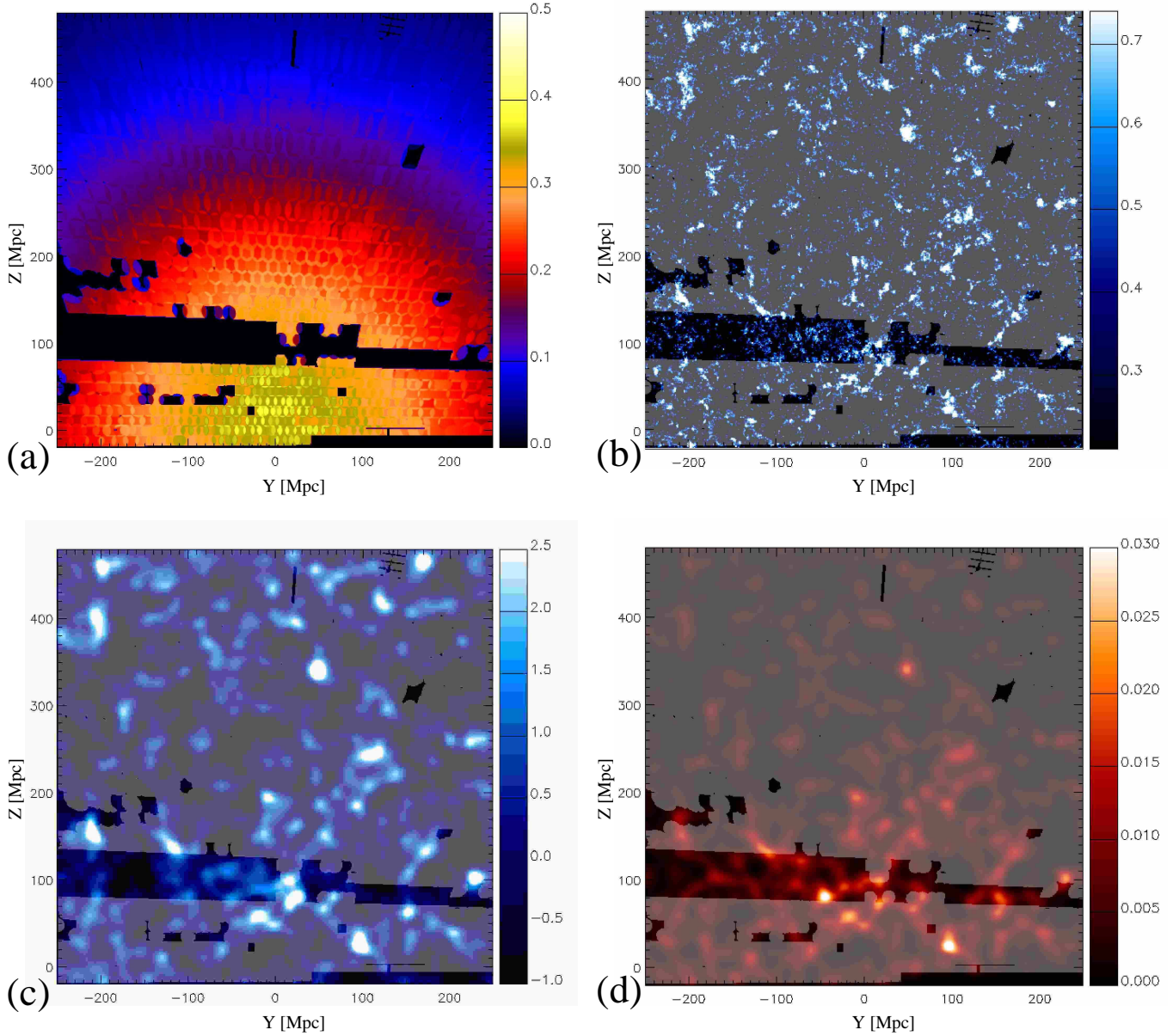[16] Note, that for a Gaussian distribution: μ_4/σ^4=3 and thereby: $k = 0$.

Figure 10. Panel (a): slice through the three dimensional mask multiplied with the selection function at $\sim$-256 Mpc in the X-axis. Panels (b) and (c) show slices through the reconstruction after taking the mean over 20 neighboring slices around the slice at $\sim$-256 Mpc in the X-axis, without smoothing and convolved with a Gaussian kernel with a smoothing radius of r_S =5 Mpc, respectively. Panel (d): DR7 sample gridded with NGP and convolved with a Gaussian kernel with a smoothing radius of r_S =5 Mpc. In panels (c) and (d) the DR6 mask is over-plotted. Note, that there is some correspondance between the structures predicted in the gap from the `Sample dr6fix` and the observed galaxy distribution there in DR7. Note, that panel (b) represents $\log(1 + \delta)$, whereas panels (c), and (d) show δ.

large deviations from Gaussianity in the unobserved regions where there is almost no data constraining the result. Note, that in Figs. 11 and 12 the skewness and kurtosis are also given (skewness: s_{10}, s_{20}, s_{30}, kurtosis: k_{10}, k_{20}, k_{30}, Pearson's skewness: s_{P10}, s_{P20}, s_{P30}, with the subscript denoting the smoothing radius in Mpc). Pearson's skewness is always larger for the observed regions than for the unobserved regions after smoothing with r_S =10 and r_S =20 Mpc and all distributions show a positive skewness. The skewness and kurtosis values show that the matter distribution starts to be closely Gaussian distributed after smoothing with a radius r_S of 30 Mpc. Nevertheless, for the region $200 < r < 400$ Mpc we find

a large deviation from Gaussianty even at that scale. Large scale structures like the Sloan Great Wall can be responsible for this. Furthermore, we analyzed in great detail the matter distribution in the region $0 < r < 600$ Mpc which has better statistics. On the right panel of Fig. 12 we can see the statistics for the unobserved region. The dashed curves show the measured distributions at different scales (black: r_S =10 Mpc, red: r_S =20 Mpc, green: r_S =30 Mpc). We calculated the means and the variances for each distribution and plotted the corresponding Gaussian distributions with light dashed-dotted lines.

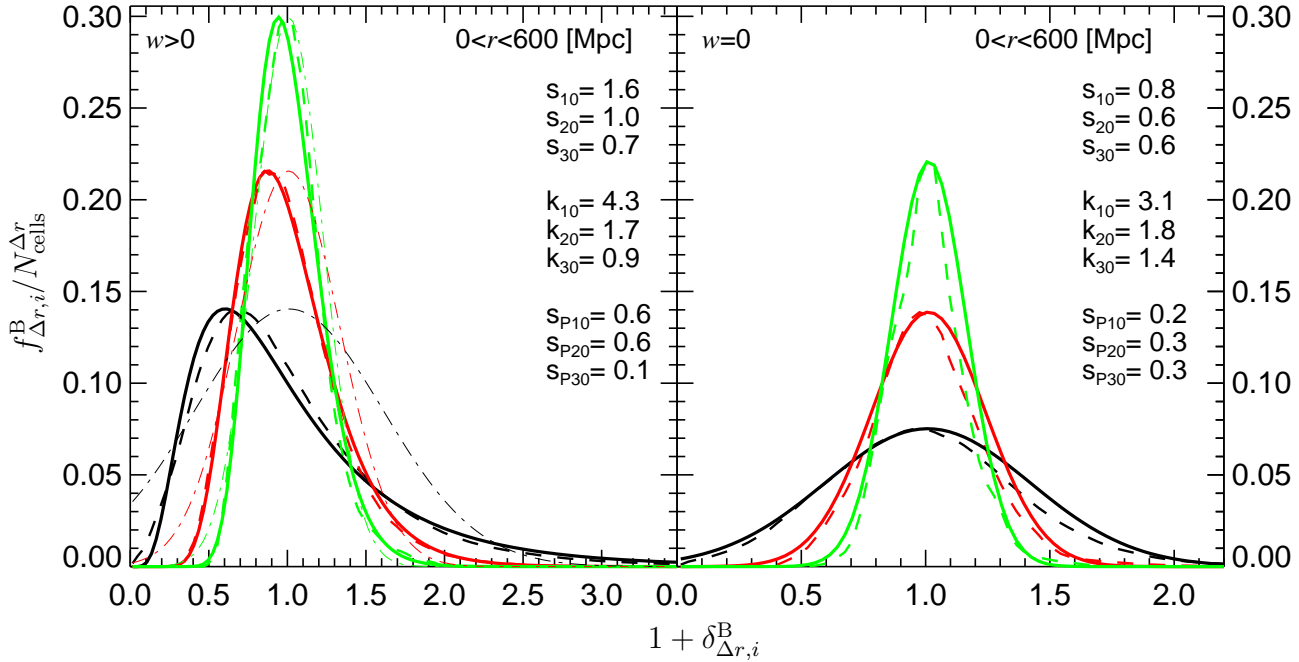On the left panel of Fig. 12 we can see the statistics for the ob-
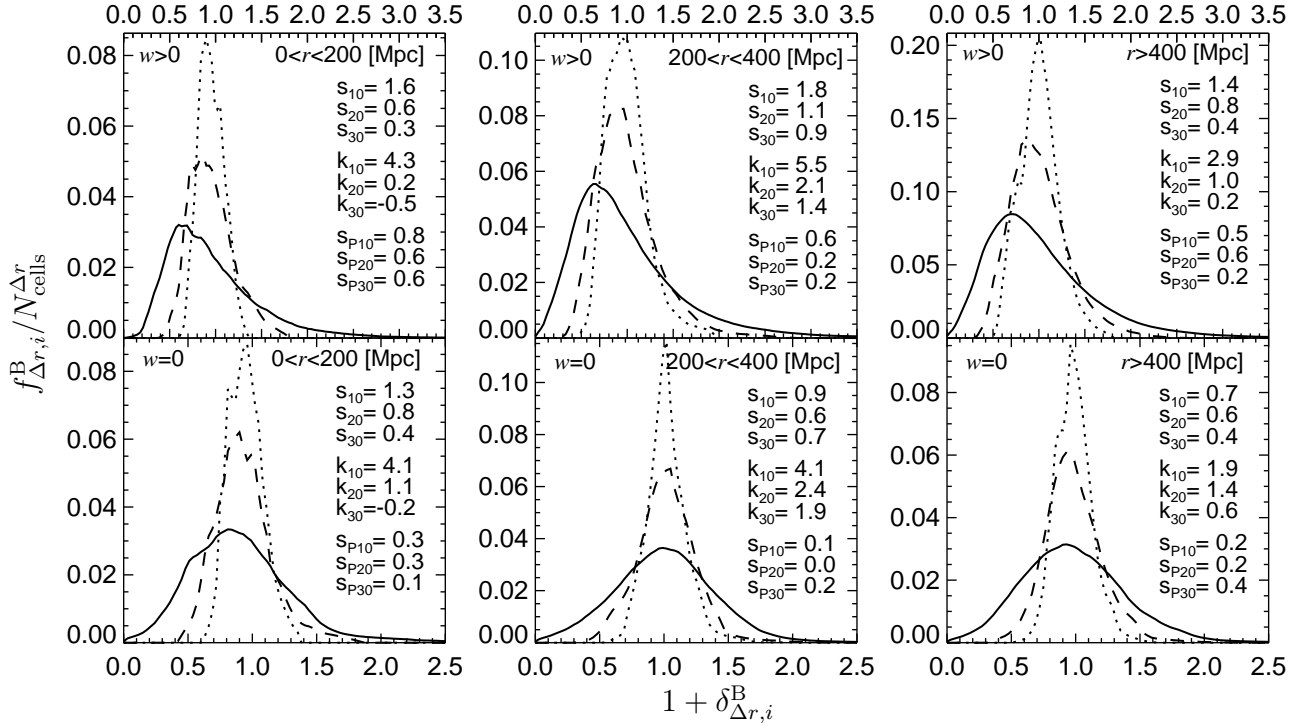
Figure 11. Statistical distribution of cells at different densities with a density binning of 0.03 in $(1 + \delta_m)$. The curves represent the distribution for the reconstructed matter field at different scales (r_S: continuous: 10 Mpc, dashed: 20 Mpc, dotted: 30 Mpc). The upper panels show the statistics at different radial shells in the observed region ($w > 0$), and the lower panels show the same in the unobserved region ($w = 0$). The corresponding skewness: s_{10}, s_{20}, s_{30}, kurtosis: k_{10}, k_{20}, k_{30}, and Pearson's skewness: s_{P10}, s_{P20}, s_{P30} are also given.

Figure 12. Statistical distribution of cells at different densities with a density binning of 0.03 in $(1 + \delta_m)$. The dashed curves represent the distribution for the reconstructed matter field at different scales (r_S: black: 10 Mpc, red: 20 Mpc, green: 30 Mpc). The corresponding skewness: s_{10}, s_{20}, s_{30}, kurtosis: k_{10}, k_{20}, k_{30}, and Pearson's skewness: s_{P10}, s_{P20}, s_{P30} are also given. On the left: (observed region: $w > 0$) continuous lines: best fit lognormal distributions using a nonlinear least squares fit based on a gradient-expansion algorithm, dashed-dotted curves: Gaussian distributions for the measured means and variances. On the right: (unobserved region: $w = 0$) continuous lines: Gaussian distributions for the measured means and variances with the corresponding statistical correlation coefficients r_{20}, r_{40}, and r_{60}.

served region with the dashed curves showing again the measured distributions at different scales (black: $r_S = 10$ Mpc, red: $r_S = 20$ Mpc, green: $r_S = 30$ Mpc). We modelled the distribution by a log-normal (see Coles & Jones 1991) and calculated the best fit using a nonlinear least squares fit based on a gradient-expansion algorithm[17]. For that, we parameterized the log-normal distribution as:

$$P(\delta_m | \boldsymbol{p}) = \frac{a}{\log(1 + \delta_m)} \exp\left[b\left(\log(1 + \delta_m) - c\right)^2\right], \quad (57)$$

with $\boldsymbol{p} = [a, b, c]$ being a set of parameters. The results of the best fits normalized with the number of cells are shown as the continuous lines on the left panel in Fig. 12. One can appreciate in all curves for $w > 0$ small tails towards low densities and long tails towards high densities showing a clear deviation from Gaussianity. The measured distributions are well fitted by the log-normal distribution of smoothed density for smoothing radii r_S of 10, 20, and 30 Mpc. We also calculated the mean and the variance and plotted the corresponding Gaussian distributions with light dashed-dotted lines. We conclude therefore, that the distribution of the matter field is in good agreement with the log-normal distribution at least in the scale range from about $10\,\mathrm{Mpc} \lesssim r_S \lesssim 30\,\mathrm{Mpc}$. This result is especially strong, since we did not assume a log-normal prior distribution in the reconstruction method. From a frequentist approach the Wiener-filter just gives the least squares estimator without imposing any statistical distribution to the matter distribution. The picture from a Bayesian perspective is more precise: a Gaussian prior distribution for the underlying density field is assumed. The posterior distribution, however, is conditioned on the data, which finally imposes its statistical behavior onto the reconstruction, as can be seen in our results.

6   CONCLUSIONS

We have presented the first application of the ARGO computer code to observational data. In particular, we have performed a reconstruction of the density field based on data from `Sample dr6fix` of the New York University Value Added Catalogue (NYU-VAGC) (see section 2). This yielded the largest Wiener-reconstruction of the Large-Scale Structure made to date requiring the effective inversion of a matrix with about $10^8 \times 10^8$ entries. The use of optimized iterative inversion schemes within an operator formalism (see Kitaura & Enßlin 2008), together with a careful treatment of aliasing effects (see Jasche et al. 2009) permitted us to recover the field on a Mpc mesh with an effective resolution of the order of ~ 10 Mpc. Furthermore, we have investigated in detail the statistical problem in particular the noise covariance employed for performing Wiener-reconstructions.

We have demonstrated that Wiener-filtering leads to different results than those obtained by the commonly used method of inverse weighting the galaxies with the selection function. Both methods are comparable when the galaxy number counts per cell is high. However, in regions with sparse observed galaxy densities inverse weighting delivers very noisy reconstructions. This finding could have important consequences in power-spectrum estimation and galaxy biasing estimation on large scales.

As part of the results the Sloan Great Wall has been presented in detail (see section 5.2) and some other prominent structures like the Coma, the Leo, and the Hercules Cluster, have been discussed, as well as the detection of a large void region (see section 5.3).

---

[17] CURVEFIT from IDL

Our results also show the detection of overdensity regions close to edges of the mask and predictions for structures in within gaps in the mask which compare well with the DR7 data in which the gaps are filled (see section 5.4). Finally, we have analyzed the statistical distribution of the density field finding a good agreement with the log-normal distribution for Gaussian smoothing with radii in the range $10\,\mathrm{Mpc} \lesssim r_S \lesssim 30\,\mathrm{Mpc}$. We hope that this work highlights the potential of Bayesian large-scale structure reconstructions for cosmology and is helpful in establishing them as a widely used technique.

REFERENCES

Abazajian K. N., Adelman-McCarthy J. K., Agüeros M. A., Allam S. S., Allende Prieto C., An D., Anderson K. S. J., Anderson S. F., et al., 2009, .Rev.Astrn.Astrophys., 182, 543

Abell G. O., Corwin Jr. H. G., Olowin R. P., 1989, .Rev.Astrn.Astrophys., 70, 1

Adelman-McCarthy J. K., Agüeros M. A., Allam S. S., Allende Prieto C., Anderson K. S. J., Anderson S. F., Annis J., Bahcall N. A., et al., 2008, .Rev.Astrn.Astrophys., 175, 297

Ballinger W. E., Peacock J. A., Heavens A. F., 1996, MNRAS, 282, 877

Bernardeau F., van de Weygaert R., 1996, MNRAS, 279, 693

Blanton M. R., Brinkmann J., Csabai I., Doi M., Eisenstein D., Fukugita M., Gunn J. E., Hogg D. W., et al., 2003a, AJ, 125, 2348

Blanton M. R., Hogg D. W., Bahcall N. A., Brinkmann J., Britton M., Connolly A. J., Csabai I., Fukugita M., et al., 2003b, ApJ, 592, 819

Blanton M. R., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I., Loveday J., 2003c, AJ, 125, 2276

Blanton M. R., Schlegel D. J., Strauss M. A., Brinkmann J., Finkbeiner D., Fukugita M., Gunn J. E., Hogg D. W., et al., 2005, AJ, 129, 2562

Carter D., Goudfrooij P., Mobasher B., Ferguson H. C., Puzia T. H., Aguerri A. L., Balcells M., Batcheldor D., et al., 2008, .Rev.Astrn.Astrophys., 176, 424

Coles P., Jones B., 1991, MNRAS, 248, 1

Cui W., Liu L., Yang X., Wang Y., Feng L., Springel V., 2008, ApJ, 687, 738

Davis M., Gerke B. F., Newman J. A., the Deep2 Team, 2005, in Astronomical Society of the Pacific Conference Series, Vol. 339, Observing Dark Energy, Wolff S. C., Lauer T. R., eds., pp. 128–+

De Lucia G., Blaizot J., 2007, MNRAS, 375, 2

Deng X.-F., Chen Y.-Q., Zhang Q., He J.-Z., 2006, Chinese Journal of Astronomy and Astrophysics, 6, 35

Doroshkevich A. G., Gottlober S., Madsen S., 1997, Astr.Astrophy.Suppl., 123, 495

Ebeling H., Wiedenmann G., 1993, Phys. Rev. E, 47, 704

Eisenstein D. J., 2005, New Astronomy Review, 49, 360

Eisenstein D. J., Annis J., Gunn J. E., Szalay A. S., Connolly A. J., Nichol R. C., Bahcall N. A., Bernardi M., et al., 2001, AJ, 122, 2267

Erdoğdu P., Lahav O., Huchra J. P., Colless M., Cutri R. M., Falco E., George T., Jarrett T., et al., 2006, MNRAS, 373, 45

Erdoğdu P., Lahav O., Zaroubi S., Efstathiou G., Moody S., Peacock J. A., Colless M., Baldry I. K., et al., 2004, MNRAS, 352, 939

Eriksen H. K., Huey G., Saha R., Hansen F. K., Dick J., Banday A. J., Górski K. M., Jain P., Jewell J. B., Knox L., Larson D. L., O'Dwyer I. J., Souradeep T., Wandelt B. D., 2007, ApJ, 656, 641

Fisher K. B., Lahav O., Hoffman Y., Lynden-Bell D., Zaroubi S., 1995, MNRAS, 272, 885

Fisher K. B., Scharf C. A., Lahav O., 1994, MNRAS, 266, 219

Frommert M., Enßlin T. A., Kitaura F. S., 2008, MNRAS, 391, 1315

Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748

Geller M. J., Huchra J. P., 1989, Science, 246, 897

Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759

Gott J. R. I., Jurić M., Schlegel D., Hoyle F., Vogeley M., Tegmark M., Bahcall N., Brinkmann J., 2005, ApJ, 624, 463

Gunn J. E., Carr M., Rockosi C., Sekiguchi M., Berry K., Elms B., de Haas E., Ivezić Ž., et al., 1998, AJ, 116, 3040

Gunn J. E., Siegmund W. A., Mannery E. J., Owen R. E., Hull C. L., Leger R. F., Carey L. N., Knapp G. R., et al., 2006, AJ, 131, 2332

Hamann J., Hannestad S., Melchiorri A., Wong Y. Y. Y., 2008, Journal of Cosmology and Astro-Particle Physics, 7, 17

Hockney R. W., Eastwood J. W., 1981, Computer Simulation Using Particles. Computer Simulation Using Particles, New York: McGraw-Hill, 1981

Hoffman Y., 1994, in ASP Conf. Ser. 67: Unveiling Large-Scale Structures Behind the Milky Way, Balkowski C., Kraan-Korteweg R. C., eds., pp. 185–+

Hogg D. W., Finkbeiner D. P., Schlegel D. J., Gunn J. E., 2001, AJ, 122, 2129

Icke V., van de Weygaert R., 1991, QJRAS, 32, 85

Ivezić Ž., Lupton R. H., Schlegel D., Boroski B., Adelman-McCarthy J., Yanny B., Kent S., Stoughton C., et al., 2004, Astronomische Nachrichten, 325, 583

Jasche J., Kitaura F. S., Ensslin T. A., 2009, ArXiv e-prints

Jewell J., Levin S., Anderson C. H., 2004, ApJ, 609, 1

Jing Y. P., 2005, ApJ, 620, 559

Jing Y. P., Börner G., 2004, ApJ, 617, 782

Jing Y. P., Mo H. J., Boerner G., 1998, ApJ, 494, 1

Kaiser N., 1987, MNRAS, 227, 1

Kim R., Strauss M., Bahcall N., Gunn J. E., Lupton R. H., Vogeley M. S., Schlegel D., the SDSS Collaboration, 2000, in Astronomical Society of the Pacific Conference Series, Vol. 200, Clustering at High Redshift, Mazure A., Le Fèvre O., Le Brun V., eds., pp. 422–+

Kitaura F. S., Enßlin T. A., 2008, MNRAS, 389, 497

Kullback S., Leibler R. A., 1951, Annals of Mathematical Statistics, 22, 79

Lahav O., 1994, in ASP Conf. Ser. 67: Unveiling Large-Scale Structures Behind the Milky Way, Balkowski C., Kraan-Korteweg R. C., eds., pp. 171–+

Lahav O., Fisher K. B., Hoffman Y., Scharf C. A., Zaroubi S., 1994, ApJ, 423, L93+

Lahav O., Lilje P. B., Primack J. R., Rees M. J., 1991, MNRAS, 251, 128

Lee J., Lee B., 2008, ApJ, 688, 78

Lee J., Li C., 2008, ArXiv e-prints

Li C., Jing Y. P., Kauffmann G., Börner G., White S. D. M., Cheng F. Z., 2006a, MNRAS, 368, 37

Li C., Kauffmann G., Jing Y. P., White S. D. M., Börner G., Cheng F. Z., 2006b, MNRAS, 368, 21

Lupton R., Gunn J. E., Ivezić Z., Knapp G. R., Kent S., 2001, in Astronomical Society of the Pacific Conference Series, Vol. 238, Astronomical Data Analysis Software and Systems X, Harnden Jr. F. R., Primini F. A., Payne H. E., eds., pp. 269–+

Mathis H., Lemson G., Springel V., Kauffmann G., White S. D. M., Eldar A., Dekel A., 2002, MNRAS, 333, 739

Meurs E. J. A., Wilkinson M. I., 1999, in Astronomical Society of the Pacific Conference Series, Vol. 176, Observational Cosmology: The Development of Galaxy Systems, Giuricin G., Mezzetti M., Salucci P., eds., pp. 333–+

Nusser A., Dekel A., 1992, ApJ, 391, 443

Panko E., Flin P., 2004, in IAU Colloq. 195: Outskirts of Galaxy Clusters: Intense Life in the Suburbs, Diaferio A., ed., pp. 245–247

Pier J. R., Munn J. A., Hindsley R. B., Hennessy G. S., Kent S. M., Lupton R. H., Ivezić Ž., 2003, AJ, 125, 1559

Ramella M., Boschin W., Fadda D., Nonino M., 2001, Astr.Astrophy., 368, 776

Richards G. T., Fan X., Newberg H. J., Strauss M. A., Vanden Berk D. E., Schneider D. P., Yanny B., Boucher A., et al., 2002, AJ, 123, 2945

Rybicki G. B., Press W. H., 1992, ApJ, 398, 169

Schaap W. E., van de Weygaert R., 2000, Astr.Astrophy., 363, L29

Schlegel D. J., Blanton M., Eisenstein D., Gillespie B., Gunn J., Harding P., McDonald P., Nichol R., Padmanabhan N., Percival W., et al., 2007, in Bulletin of the American Astronomical Society, Vol. 38, Bulletin of the American Astronomical Society, pp. 966–+

Schmoldt I. M., Saar V., Saha P., Branchini E., Efstathiou G. P., Frenk C. S., Keeble O., Maddox S., McMahon R., Oliver S., Rowan-Robinson M., Saunders W., Sutherland W. J., Tadros H., White S. D. M., 1999, ApJ, 118, 1146

Sheth J. V., Sahni V., 2005, ArXiv Astrophysics e-prints

Smith J. A., Tucker D. L., Kent S., Richmond M. W., Fukugita M., Ichikawa T., Ichikawa S.-i., Jorgensen A. M., et al., 2002, AJ, 123, 2121

Smith R. E., Peacock J. A., Jenkins A., White S. D. M., Frenk C. S., Pearce F. R., Thomas P. A., Efstathiou G., Couchman H. M. P., 2003, MNRAS, 341, 1311

Spergel D. N., Bean R., Doré O., Nolta M. R., Bennett C. L., Dunkley J., Hinshaw G., Jarosik N., et al., 2007, .Rev.Astrn.Astrophys., 170, 377

Springel V., White S. D. M., Jenkins A., Frenk C. S., Yoshida N., Gao L., Navarro J., Thacker R., Croton D., Helly J., Peacock J. A., Cole S., Thomas P., Couchman H., Evrard A., Colberg J., Pearce F., 2005, Nature, 435, 629

Stoughton C., Lupton R. H., Bernardi M., Blanton M. R., Burles S., Castander F. J., Connolly A. J., Eisenstein D. J., et al., 2002, AJ, 123, 485

Strauss M. A., Weinberg D. H., Lupton R. H., Narayanan V. K., Annis J., Bernardi M., Blanton M., Burles S., et al.,, 2002, AJ, 124, 1810

Struble M. F., Rood H. J., 1999, .Rev.Astrn.Astrophys., 125, 35

SubbaRao M., Frieman J., Bernardi M., Loveday J., Nichol B., Castander F., Meiksin A., 2002, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Starck J.-L., Murtagh F. D., eds., Vol. 4847, pp. 452–460

Tegmark M., Blanton M. R., Strauss M. A., Hoyle F., Schlegel D., Scoccimarro R., Vogeley M. S., Weinberg D. H., et al.,, 2004, ApJ, 606, 702

Thomsen B., Baum W. A., Hammergren M., Worthey G., 1997, ApJ, 483, L37+

Tucker D. L., Kent S., Richmond M. W., Annis J., Smith J. A., Allam S. S., Rodgers C. T., Stute J. L., et al.,, 2006, Astronomische Nachrichten, 327, 821

van de Weygaert R., Schaap W., 2001, in Mining the Sky, Banday A. J., Zaroubi S., Bartelmann M., eds., pp. 268–+

Wandelt B. D., Larson D. L., Lakshminarayanan A., 2004, Phys. Rev. D, 70, 083511

Webster M., Lahav O., Fisher K., 1997, MNRAS, 287, 425

Wiener N., 1949, Extrapolation, Interpolation, and Smoothing of Stationary Time Series. New York: Wiley

York D. G., Adelman J., Anderson Jr. J. E., Anderson S. F., Annis J., Bahcall N. A., Bakken J. A., Barkhouser R., et al.,, 2000, AJ, 120, 1579

Zaninetti L., 1995, Astr.Astrophy.Suppl., 109, 71

—, 2006, Chinese Journal of Astronomy and Astrophysics, 6, 387

Zaroubi S., Hoffman Y., Dekel A., 1999, ApJ, 520, 413

Zaroubi S., Hoffman Y., Fisher K. B., Lahav O., 1995, ApJ, 449, 446

Zel'dovich Y. B., 1970, Astr.Astrophy., 5, 84